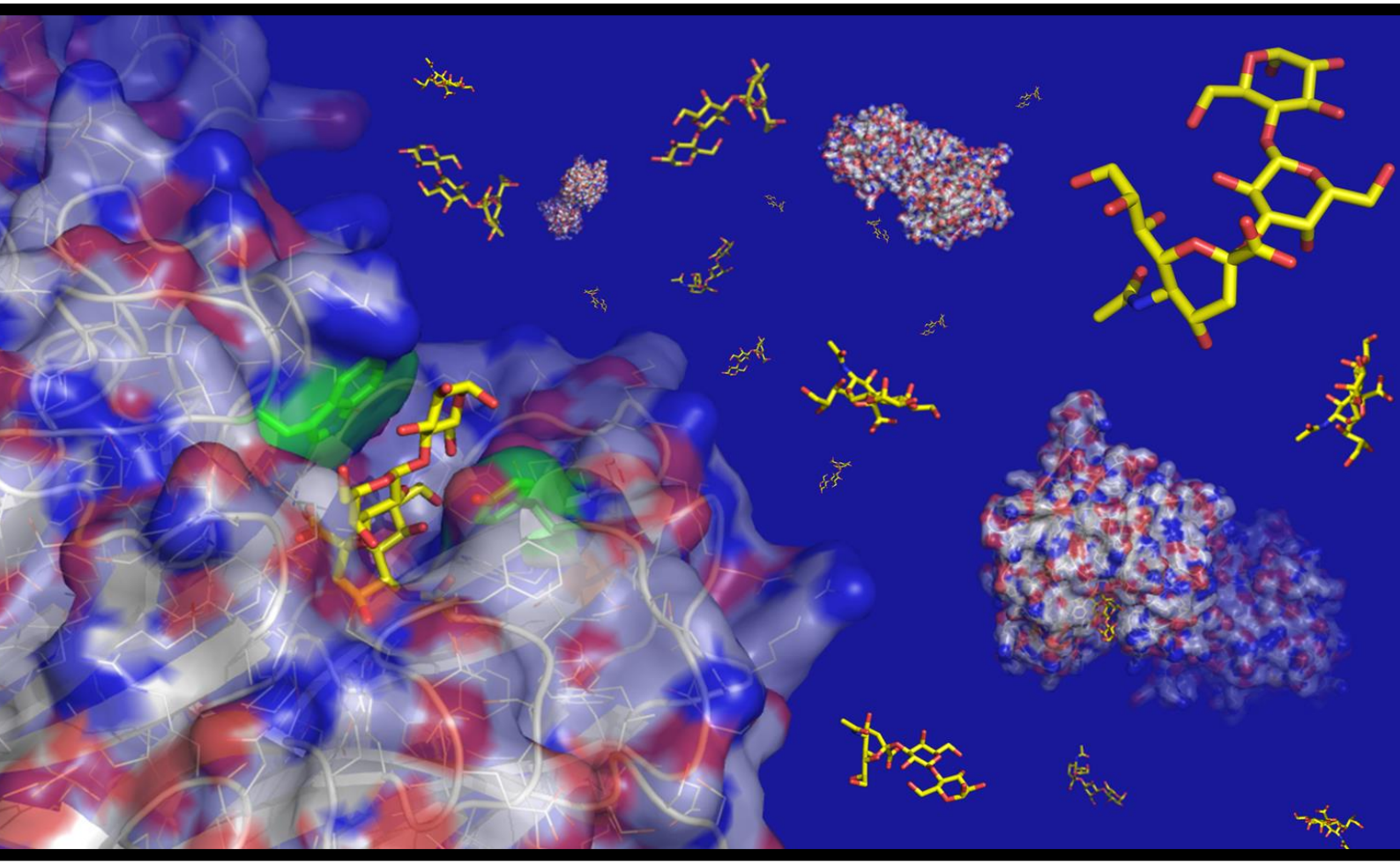
Production of prebiotic oligosaccharides by novel enzymatic catalysis

Nordvang, Rune Thorbjørn

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Production of prebiotic oligosaccharides by novel enzymatic catalysis



Rune Thorbjørn Nordvang

PhD thesis

2015

Preface

The work presented in the this thesis was conducted during my PhD project at the Center of Bioprocess Engineering (BioEng), Department of Chemical and Biochemical Engineering at the Technical University of Denmark (DTU), between December 2012 and December 2015.

The work was funded by *Mejeriernes Forsknings Fond* (MFF) and through a DTU stipend. The work was carried out as part of a larger project with participation of several academic research groups as well as researchers from industry. The project was intentionally designed to cover as many aspects of the overall project within the scope of enzymatic production of the relevant prebiotic oligosaccharides. Thus is was very appropriate that the main supervisor for my PhD project was the project leader of the overall project, Jørn Dalgaard Mikkelsen (BioEng) but due to the broad nature of the project supervision was given by several co-supervisors Manuel Pinelo and Carsten Jers.

Summary

A group of prebiotic oligosaccharides known as human milk oligo-saccharides (HMOs) are currently receiving a lot of attention due to the prospect of their addition to infant formula. Whereas prebiotics in general are used as mediators for modulating the gut microbiome in human individuals, HMOs play an important role in development of this organ, where it contributes to the selective growth stimulation of the beneficial microorganism *Bifidobacterium infantis*. The effects of HMOs are not only prebiotic and a range of beneficial effects have been postulated, with varying amounts of scientific evidence backing them up.

Since chemical synthesis of carbohydrates is extremely cumbersome, it is generally accepted that HMOs must be produced biochemically and enzymatic in vitro production is a popular strategy. Thus, the purpose of this PhD project was to encompass as many of the aspects of the enzymatic production of HMOs as possible, and identify opportunities to improve the enzymes, reaction efficiencies and processes involved.

For enzymatic in vitro production of HMOs, industrial side stream products are often used as substrates to reduce the final product price. However, to use these substrates it is generally necessary to identify glycosyl hydrolases with trans-glycosidase activity or ideally rare trans-glycosidases. The BioEng group has previously developed a state of the art engineered trans-sialidase used for the synthesis of sialylated HMOs. Thus, synthesis of the simple genuine mono-sialylated HMO, 3'sialyllactose(3'SL), received particular attention in this PhD project. The BioEng state of the art trans-sialidase was, during this PhD project, further mutated, raising the bar for competing enzymes. For further improvement of the current leading enzyme, it was concluded that new knowledge would be required and that such knowledge could be provided by identification of novel trans-sialidases, which have, however, only been identified in a single genus. Never the less, as part of this PhD project a novel trans-sialidase was identified which was capable of producing 3'SL and a novel trans-sialylation product, 3SL, the properties of which are unknown.

With the goal to further improve 3'SL production, the process strategy underwent scrutiny and weak points were identified and improved upon. At the start of the PhD project, 3'SL was purified in a three step process including ultrafiltration, with subsequent column chromatography and removal of eluent. As part of this PhD project, an innovative nanofiltration approach eliminated the necessity for column chromatography and eluent removal. Furthermore, by moving the HMO enzymatic synthesis to a membrane reactor, an integrated membrane system strategy was constructed and proof of concept was demonstrated.

From the beginning of the PhD project, it was known that future endeavors would include the synthesis of larger HMO structures, for which enzymes and substrates for HMO backbone synthesis would be required. Progress in this aspect of HMO production was also achieved during this PhD project, as two novel β-N-acetylhexosaminidases were identified through screening of metagenomic libraries. Both enzymes were successfully used to produce HMO backbone precursors, which have previously been used for HMO backbone synthesis.

Resumé

For tiden bliver der forsket meget i en bestemt gruppe af prebiotiske oligosaccharider kaldet HMO'er (mælke-oligosaccharider af human oprindelse), med det formål at fremstille og tilsætte dem til modermælkserstatning. Hvor prebiotika generelt anvendes til at modulere på et individs tarmflora, er HMO'erne ekstra interessante fordi de er stærkt medvirkende til at udvikle den første tarmflora der etableres i et spædbarn. Dette sker bl.a. igennem en stimulering af væksten af *Bifidobacterium infantis*, som er en gavnlig bakterie. HMO'er menes, foruden at være prebiotiske, at have en række gavnlige egenskaber som, i varierende grad, kan bakkes op videnskabeligt.

Da kemisk syntese af kulhydrater er besværligt er der generelt en opfattelse af at det er nødvendigt at anvende biokemisk syntese til HMO-fremstilling og *in vitro* enzymatisk fremstilling er en populær metode. Det var således formålet med dette PhD-projekt at omfatte så mange aspekter af enzymatisk HMO production som muligt, med det formål at identificere muligheder for at forbedre enzymer, reaktionsbetingelser og processer.

Til *in vitro* produktion af HMO'er bruges industrielle bi-produkter ofte som substrater for at reducere produktionsomkostningerne. For at anvende disse substrater er det, til gengæld ofte nødvendigt at tilvejebringe hydrolytiske enzymer med trans-glycosidaseaktivitet eller ideelt de sjældne trans-glycosidaser. Forskningsgruppen, BioEng, har tidligere udviklet en gen-manipuleret trans-sialidase der, som en del af dette projekt, blev videreudviklet og har sat barren endnu højre for konkkurerende enzymer. For at videreudvikle på dette enzym var der enighed om at ny viden ville være en nødvendighed og at sådan viden kunne opnås hvis en ny trans-sialidase kunne identificeres. Trans-sialidaser er dog, indtil videre, kun identificeret i en enkelt slægt. Ikke desto mindre blev en ny trans-sialidase identificeret som en del af dette PhD-projekt og det nye enzym var i stand til at syntetisere 3'SL samt et nyt trans-sialyleringsprodukt, 3SL, hvis egenskaber er ukendte.

Med det mål at forbedre 3'SL fremstilling blev den anvendte proces taget nøje I betragtning, så svage led kunne identificeres og forbedres. Ved PhD-projektets begyndelse udførtes oprensning af 3'SL i en tretrins proces: Ultrafiltrering efterfulgt af kolonne-oprensning og separation af eluent. Som del af dette PhD-projekt blev kolonne-oprensning og eluent-separation erstattet af en innovativ anvendelse af nano-filtrering. Endvidere blev et integreret membran-system udviklet og anvendt til produktion af 3'SL.

Fra PhD-projektets begyndelse var det klart at det I fremtiden vil være målet at syntetisere større HMO strukturer. For at dette kan lade sig gøre vil det være nødvendigt at identificere enzymer og substrater til formålet. Dette PhD-projekt bidrog også med fremgang i denne henseende idet to nye β-N-acetylhexosaminidaser blev identificeret ved hjælp af et DNA-bibliotek fremstillet af oprenset DNA fra en jordprøve. Begge enzymer blev med succes anvendt til at fremstille molekyler der er forstadier til HMO-backbones og som tidligere er anvendt til at fremstille disse.

Acknowledgements

Firstly, I would like to thank Jørn Dalgaard Mikkelsen, Manuel Pinelo, Carsten Jers and Center Director Anne S. Meyer for their invaluable input to my research, for the scientific discussions, and for letting me take charge of the research strategies when my ideas were not too ambitious.

For their contributions to my research I would also like to thank; most importantly Christian Nyffenegger and Jian Quan Luo, but also Birgitte Zeuner, Rasmus Prior, Mads Friis Andersen, Sofie Tage Mortensen, Jesper Holck, Matheus Łężyk, Nanna Villumsen, Ulrik Kræmer Sundekilde, Elisabetta Difilippo, Madelon Lochtenberg, and Henk Schols.

I would like to thank Nanna Villumsen and Malwina Michalak for the opportunity to collaborate on interesting research topics and OMICS international for the opportunities to present my work overseas in both presentation and poster format.

Of course my colleagues at the Department of Chemical and Biochemical Engineering, where especially those at BioEng, deserve much gratitude. The same goes for the DTU Exiles R.U.F.C. – both colleagues and team mates were strong contributing factors for my decision to pursue a PhD degree at DTU.

Finally, I cannot express how much I want to thank you; my beautiful wife, Emily Catherine Nordvang and the little pink terror, my daughter Erica Rose Nordvang, for your support and the motivation you have given me (one of you, for your whole life) throughout this test. The two of you can always make me smile!

Rune Thorbjørn Nordvang
Kgs. Lyngby
December 2015

List of abbreviations

3SL	3-sialyllactose
3'SL	3'-sialyllactose
6'SL	6'-sialyllactose
AA	Amino acid
CGMP	Casein glycomacropeptide
CI	Covalent intermediate
EMR	Enzymatic membrane reactor
FOS	Fructo-oligosaccharides
GH	Glycoside hydrolase
GlcNAc	N-acetyl-glucosamine
GOS	Galacto-oligosaccharides
GSSM	Gene Site Saturation Mutagenesis™
HMO	Human milk oligosaccharides
HPAEC-PAD	High performance anion exchange chromatography – pulsed amperimetric detection
LNnT	Lacto-N-neotetraose
LNT	Lacto-N-tetraose
LNT2	Lacto-N-triose
MALDI-TOF	Matrix assisted laser dissociation ionization – time of flight
MWCO	Molecular weight cutoff
NF	Nanofiltration
NMR	Nuclear magnetic resonance
pNP	para-Nitrophenol
PKU	Phenylketonuria
RC	Regenerated Cellulose
SA	Sialic acid
TcTS	*Trypanosoma cruzi* trans-sialidase
TrSA	*Trypanosoma rangeli* sialidase
UF	Ultrafiltration

Contents

List of Publications

Included in this thesis:

Paper I (in preparation): Nordvang RT, Nyffenegger C, Holck J, Jers C, Zeuner B, Sundekilde UK, Meyer AS, Mikkelsen, JD (2015), *It all starts with a sandwich - A rational 3D alignment approach to the identification of sialidases with trans-glycosylation activity*

Paper II: Nyffenegger C, **Nordvang RT**, Zeuner B, Łężyk M, Difilippo E, Logtenberg MJ, Schols HA, Meyer AS, Mikkelsen JD (2015), *Backbone structures in human milk oligosaccharides: trans-glycosylation by metagenomic β-N-acetylhexosaminidases*

Paper III (in preparation): Nyffenegger C**, Nordvang RT**, Jers C, Meyer AS, Mikkelsen JD (2015), *Enhancing trans-sialidase activity in a Trypanosoma rangeli sialidase by site-directed mutagenesis*

Paper IV: Luo J, **Nordvang RT**, Morthensen ST, Birgitte Zeuner, Meyer AS, Mikkelsen JD, Pinelo M (2014), *An integrated membrane system for the biocatalytic production of 3 '-sialyllactose from dairy by-products*

Paper V: Nordvang RT, Luo J, Zeuner B, Prior R, Andersen MF, Mikkelsen JD, Meyer AS, Pinelo M (2014), *Separation of 3'-sialyllactose and lactose by nanofiltration: A trade-off between charge repulsion and pore swelling induced by high pH*

Not included in this thesis:

Villumsen NS, Jensen HB, Le TTT, Møller HS, **Nordvang RT**, Nielsen LR, Nielsen SB, Sørensen J, Hammershøj M, Larsen LB (2015), *Self-assembly of caseinomacropeptide as a potential key mechanism in the formation of visible storage induced aggregates in acidic whey protein isolate dispersions*

Michalak M, Larsen DM, Jers C, Almeida JRM, Willer M, Li H, Kirpekar F, Kjærulff L, Gotfredsen CH, **Nordvang RT**, Meyer AS, Mikkelsen JD (2013), *Biocatalytic production of 3'-sialyllactose by use of a modified sialidase with superior trans-sialidase activity*

Chapter 1: Introduction

This PhD project is rooted in the industrial application of enzymes for the production of human milk oligo-saccharides (HMOs) and throughout the thesis industrial application has been kept in mind, thus the reported PhD project has strived to live up to the DTU mantra *"Det blir' til noget",* which for unimaginable reasons is written in Danish, but roughly translates to: "something becomes of it" or translated in essence: "real life solutions".

Goals

- Identification of novel enzymes applicable for production of genuine HMOs
- Improvement of in house enzymes for production of HMOs
- Optimization of HMO production with special attention to scalability, purification and substrate utilization.

Thesis structure

In the reminder of this Chapter, key aspects of enzymatic production of prebiotics will be introduced. Upon reaching the end of Chapter 1 you, the reader of this thesis should be prepared to appreciate the proposed hypotheses and be invigorated to dive into the scientific work carried out as part of this PhD project. In each of the Chapters 2-6, the publications will be preceded by four parts; motivation, hypotheses and objectives, experimental considerations, and conclusion, to put the article into the context of the overall PhD project. Hence, the actual publications may include additional hypotheses and conclusions (than highlighted in the Chapter introduction) because it was occasional beneficial to scientific progress to venture outside the scope of the overall PhD project. In Chapters 2 and 3, the aspects of novel enzyme discovery are covered using two different routes to identification of trans-glycosidases, and dealing with the challenges involved with identifying this ingenious class of enzymes. Chapter 4, deals with the aspects of enzyme optimization is addressed and the state of the art engineered trans-sialidase Tr13 developed by our research group was modified by a rational approach and evaluated *in vitro*. In Chapter 5, the last steps of HMO production are carried out, when an innovative application of NF is used to purify the HMO 3'SL. Several pieces of the PhD project come together, in Chapter 6, as a process for optimized 3'SL production is proposed with the concepts demonstrated. Finally, the overall conclusions of the PhD project as a whole will be drawn up in Chapter 7.

Prebiotic oligosaccharides

The term 'Prebiotic' was coined in 1956, describing molecular species prior to, but leading to life. However the term was kidnapped in 1995 Gibson and Roberfroid who thoroughly introduced the prebiotic concept and defined the term prebiotics as "non-digestible oligosaccharides that reach the colon without being hydrolysed and are selectively metabolized by health-positive bacteria such as bifidobacteria and lactobacilli thereby exerting a beneficial effect on the host health"(G. R. Gibson & Roberfroid, 1995). The concept, upon introduction, received much attention from both academia and industry and many foods were proclaimed to have prebiotic activity. As a consequence the concept was revised several times(Glenn R Gibson, Probert, Loo, Rastall, & Roberfroid, 2004; Roberfroid, 2007) and in 2007 same Roberfroid offered a refined definition of prebiotics as: "a selectively fermented ingredient that allows specific changes, both in the composition and/or activity in the gastrointestinal microflora that confers benefits upon host well-being and health"(Roberfroid, 2007). Although the definition was changed the overall concept still revolve around the growth stimulation of bifidobacteria and lactobacilli and the usual suspect for such growth stimulation continues to be oligosaccharides such as fructo-oligosaccharids (FOS), galacto-oligosaccharides (GOS) and polysaccharides such as inulin(Charalampopoulos & Rastall, 2012).

Commercially applied Prebiotics, infant formula and HMOs

Whereas some foods are inherent prebiotic due to a natural content of a prebiotic component (i.e. Jerusalem Artichokes containing large amounts of the prebiotic inulin)(Moshfegh, 1999) a variety of commercially available products have prebiotics added to them in an effort to improve the health benefits of the product(Charalampopoulos & Rastall, 2012). One of the most pronounced uses of prebiotics is in the preparation of infant formula(Sabater, Prodanov, Olano, Corzo, & Montilla, 2016), which is also the main topic of this PhD thesis. Human milk contains, besides lactose, a tremendous amount of oligosaccharides collectively designated as HMOs(Bode, 2012). The complexity and nature of the HMOs will be further introduced, but among other beneficial effects they serve as prebiotics initializing and guiding the constitution of the infant microbiome(Clemens Kunz & Rudloff, 1993). Since the production of HMOs for infant formula is in its infancy, FOS and GOS is currently added to most infant formulas where they serve as an HMO substitute(Sabater et al., 2016). The composition of HMOs is incredibly diverse and varies from individual to individual as well as over the course of the breast feeding period as reviewed by Kunz et al.(C Kunz, Rudloff, Baier, Klein, & Strobel, 2000). MHOs specifically stimulates the growth of beneficial bacteria such as *Bifidobacterium infantis*, which is especially efficient at metabolizing HMOs as it processes a large gene cluster dedicated to HMO consumption(Sela & Mills, 2010). The prebiotic effect of HMOs are however just one of many, to the infant, beneficial effects claimed to be associated with HMOs as summarized in Fig. 1 from an excellent review on the subject by Bode(Bode, 2012).

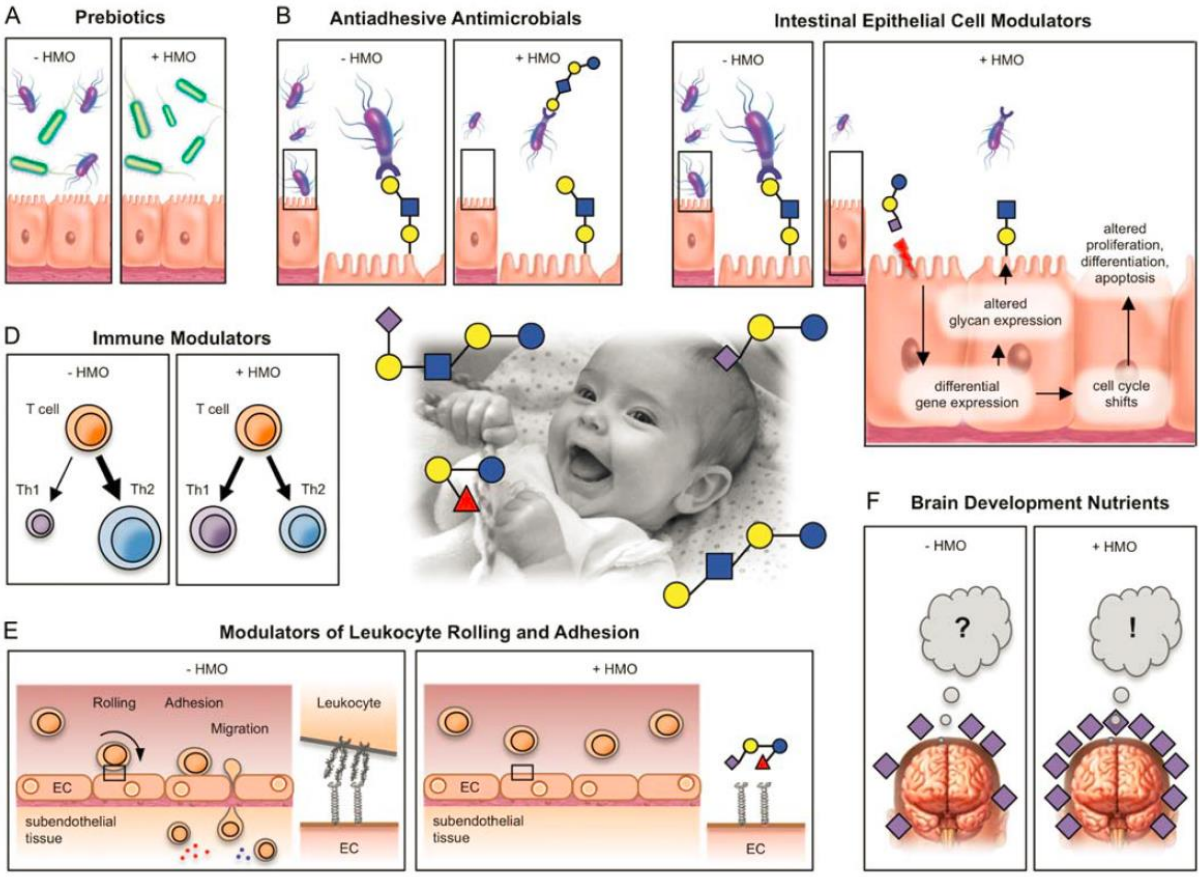


Figure 1 (adopted from Bode 2012(Bode, 2012)): Many beneficial effects have been suggested by the scientific community: (A) As prebiotics HMOs serve as energy source enabling the beneficial bifidobacteria to outcompete potential harmful microorganisms. (B) The HMOs resemble structures on the human epithelium and are thus adhere to glycan receptors of pathogen organisms preventing its attachment. (C)HMOs affect gene regulation in epithelial cells, which in turn changes the cell glycosylation pattern and other cell responses. (D) HMOs affect the lymphocyte cytokine production which is suggested to aid a healthier Th1/Th2 response. (E) HMOs inhibit the leukocyte rolling and on endothelial cells, activated by the immune system, through inhibition of selectin-mediated cell-cell interactions. (F) Sialic acid is an essential building block in brain tissue and sialylated HMOs may therefore serve as a source thereof, improving brain cognitive development.

Production of prebiotic oligosaccharides

Although some foods contain prebiotic oligosaccharides as a naturally occurring component (e.g. FOS and inulin in chicory root(Franck, 2002) of GOS in soybean(Espinosa-Martos & Rupérez)) most of oligosaccharides used for food functionalization are prepared by enzymatic reactions(Patel & Goyal, 2012). In principal there are two different strategies for obtaining oligosaccharides enzymatically – a top down and

a bottom up strategy(Patel & Goyal, 2012). In the top down approach carbohydrate polymers are partially degraded into their constitutive oligo-mers. This approach has long been used for the production of FOS from inulin(Bornet, Brouns, Tashiro, & Duvillier, 2002), but promising research apply the approach to produce compounds with prebiotic effects from more complex polymers such as pectin(Holck et al., 2011). The alternative bottom up strategy, which is the strategy investigated in this PhD study, relies on enzymes to build the oligosaccharides by addition of 1-2 glycosides at a time. The bottom up strategy is the strategy used by most living organisms for synthesis of their native polysaccharides, but whereas these reactions in living organisms are carried out by a class of enzymes known as glycosyltransferases(Ban et al., 2012) another class of enzymes are investigated in this PhD study. The enzymes studied in this PhD study all belong to classes of enzymes with hydrolytic activity towards carbohydrates, a decision that might at first seem idiotic but which have been successfully applied(Zeuner, Jers, Mikkelsen, & Meyer, 2014) and, hopefully, will seem convincing at the end of the thesis. The reason for looking away from the glycosyltransferases is due to their limitation in substrate specificity. Whereas glycosyltransferases are generally relying on nucleotide activated glycoside substrates, it is the hope that the promiscuity of hydrolases showing affinity for a range of polysaccharides will enable the use of industry side-stream products as substrates for prebiotic oligosaccharide production.

HMO synthesis – building the holy grail of prebiotics bottom up

As mentioned human breast milk contain vast amounts of HMOs. HMOs are intrinsically of human origin, but similar structures can be obtained in the breast milk from other mammalian species such as elephants and camels(C. Kunz, 1999; Urashima, Taufik, Fukuda, & Asakuma, 2013). In bovine milk however HMO-like structures are very limited (with concentrations at about a 5% of Human milk) and variation in structure from human milk reduces the applicability further(Urashima et al., 2013). Thus natural sources of HMOs are limited and thus strategies to produce them are being developed(Holck et al., 2014; Michalak et al., 2014; Zeuner, Jers, et al., 2014; Zeuner, Luo, et al., 2014) – an effort which this PhD project revolves around. Lactose is the precursor of all HMOs identified so far(Bode, 2012). The lactose can then either be elongated by N-acetyllactosamine in a β1→6 or β1→3 fashion or by lacto-N-biose in a β1→3 fashion (Bode, 2012). As if these possible elongations did not permit enough diversity the resulting branched or unbranched backbone can be further decorated with fucose and or sialic acid. Fucose is found in HMOs in an α1→2, α1→3, and an α1→4 bound fashion whereas sialic acid is found in an α2→3 or α2→6 bound fashion (Bode, 2012). The general principle of HMO structures is shown in Fig. 2, which however gives a slightly distorted picture of HMO diversity. The possible diversity of structures and associated complexity is vast and nature utilizes the possibilities to their fullest, making life harder for the scientists trying to replicate the synthesis of these structures (e.g. the author of this thesis). Thus more than 200 different HMO structures have been

identified so far(Ninonuevo et al., 2006) and thus it is extremely difficult to establish which HMOs are essential to gain each of the claimed beneficial effects of the HMOs. Among the structures shown in Fig. 2 are four molecules, which have been the target for the enzymatic reactions and processes developed in this PhD thesis: 3'Sialyllactose (3'SL), 6'-sialyllactose (6'SL), Lacto-N-tetraose (LTN) and Lacto-N-neotetraose (LNnT).

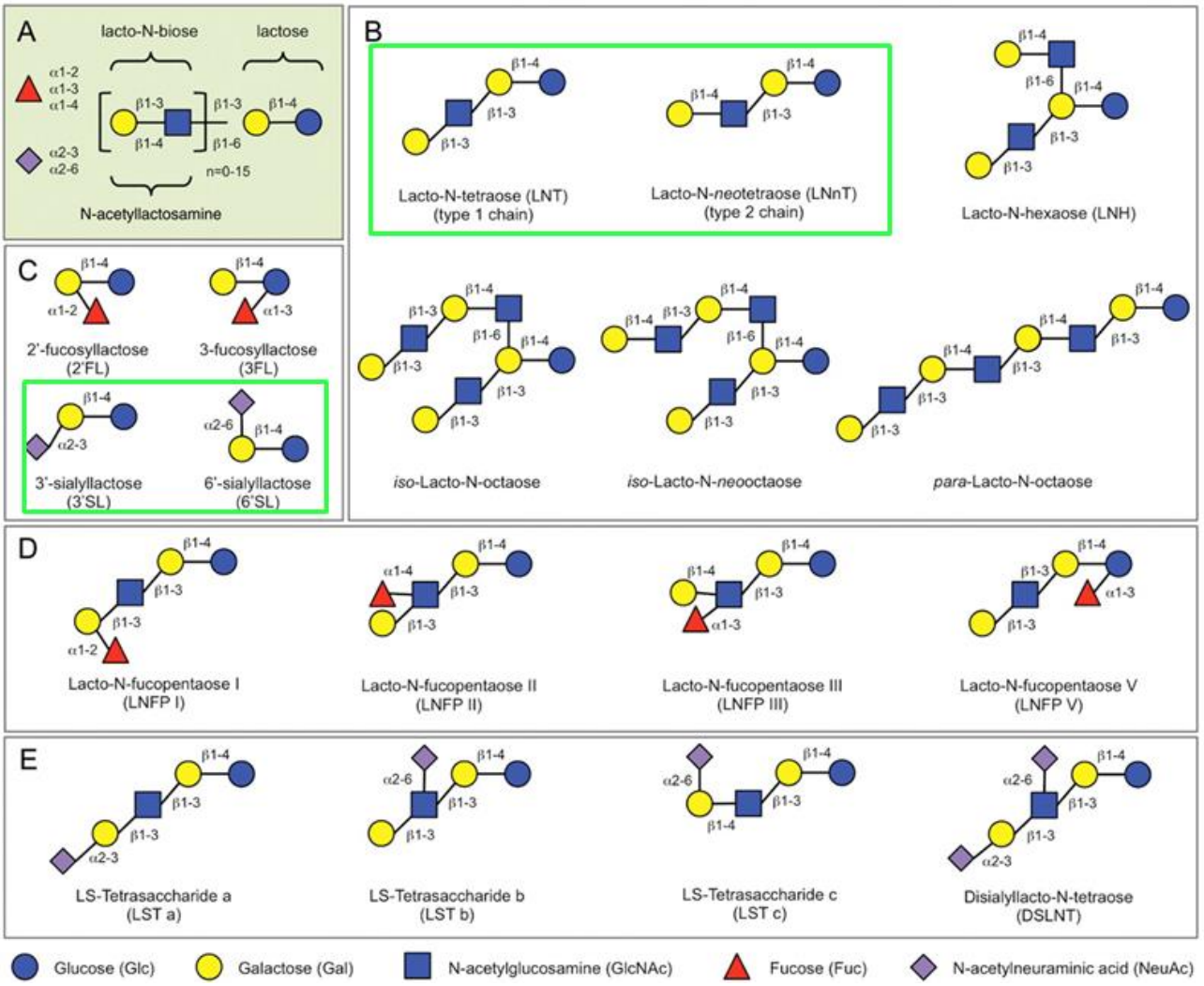


Figure 2 (adapted from Bode 2012(Bode, 2012)): HMOs are a complex mixture of oligosaccharides synthesized from up to five building blocks (A); lactose, lacto-N-biose, Nacetyllactosamine, sialic acid, and fucose. lactose act as the initial precursor for all HMOs and is recognized in all HMOs at the reducing end. A longer backbone may be generated by addition of lacto-N-biose or N-acetyllactosamine units before the HMOs may be (D) fucosylated (in an α1→2, α1→3, and an α1→4) or (E) sialylated (in an α2→3 or α2→6 fashion). The structures that are produced during the research of this PhD thesis are (green frames) 3'SL, 6'SL, LNT and LNnT.

The building blocks for HMOS

When developing the process for HMO production it is of utmost importance to be well acquainted with the substrates that are to be utilized. Substrates for the enzymatic reactions, involved in the production of HMOs should originate from non-toxic or ideally food-grade compounds to ensure that the final product is safe and appropriate for consumption. Safety is however not the only factor to take into account when deciding on the substrates for the process. As mentioned the HMOs are intended for improving infant formula which is a large bulk product in a competitive market(Kent, 2015). Although infant health is important to the parents around the world, the price must be at a level where parents can be persuaded to pay for the new product. Thusly, availability and the associated price is another key factor to take into account. Thus industrial side stream products are often taken into consideration when substrates are searched for. Furthermore, to optimize feasibility of the HMO production process, it must be evaluated whether the substrates can be used or reused for a different purpose as it will further reduce the price (due to a second revenue stream) as well as the overall sustainability of the process.

Since this PhD project revolves around the development of enzymatic HMO production in a bottom up fashion, the required substrates are: A suitable HMO backbone precursor as well as donors of sialic acid, fucose and N-acetyl glucosamine. In this regard the term donor describes a substrate, in which the sugar of interest is covalently bound and thusly has the energy needed to form the enzyme-substrate intermediate (CI) as further discussed in the section regarding enzymes for HMO production.

HMO-precursor Lactose

As lactose is the basis for all HMOs, residing at the reducing end of all identified structures, it is the natural choice for an HMO backbone precursor. Lactose is readily available in high quantities and is already used as the substrate for GOS synthesis(Wang et al., 2012). In the enzymatic reactions lactose will act as the acceptor molecule, in the transfer-glycosylation reaction mediated by the enzyme – a reaction which has also been demonstrated for 3'SL production(Holck et al., 2014).

N-acetylgalactosamine donor - Chitin

For more complex backbone molecules it is necessary to elongate the lactose precursor with either of the two disaccharides lacto-N-biose or N-acetyllactosamine. An enzymatic process transferring either of the di-saccharides directly to a lactose molecule could be imagined, but no suitable donor substrates have been identified. Thus another strategy is necessary and Murata et al. have shown that synthesis of LNT and LNnT can be carried out enzymatically in a two-step reaction adding one mono-saccharide moiety at a time(Murata, Inukai, Suzuki, Yamagishi, & Usui, n.d.). In the first reaction step, a nucleotide-activated substrate was used as the donor substrate for a β-1,3-N-acetylglucosaminyltransferase from bovine serum

to carry out the transfer of N-acetylglucosamine (GlcNAc) to the acceptor; lactose(Murata et al., n.d.). Genuine HMOs (LNT and LNnT) were synthesized by subsequent addition of a galactose (respectively β1→3 and β1→4 bound) in two different β-D-galactosidase-mediated trans-glycosylations(Murata et al., n.d.). Murata argues that the synthesizing process is suitable for large-scale operations(Murata et al., n.d.), but as previously described the nucleotide activated donors are undesired due to price and availability. Thus an alternative donor is required for HMO production and a candidate has been identified in the form of polymer chitin, which is the polymer composing the exo-skeleton of arthropods (insects, crustaceans, arachnids and myriapods)(Prabu, 2012) and have demonstrably been isolated from industrial shellfish waste (shells) (Arbia, 2013; Ghorbel-Bellaaj, Younes, Maâlej, Hajji, & Nasri, 2012). Thus, using chitin as an GlcNAc-donor valorize large amounts of industrial shrimp shells waste and could therefore be an ideal donor from an availability point of view. Another advantage of using shrimp shell is that its origin grants it food grade status. However, using shrimp shell is not without challenges as chitin is insoluble as a polymer(Arbia, 2013) (which is appreciated by the arthropods) and therefore needs preprocessing before it can be utilized for its intended purpose. Finally using chitin from this source could pose a threat to individuals with shellfish allergy.

Sialic acid donor - CGMP

Casein glycol macro peptide (CGMP) is as 64 AA peptide chain, which as the name suggests originates from the milk protein, casein, which takes up 80% of the protein in bovine milk. The peptide forms due to partial degradation of the κ-casein by chymosin during cheese production, where it end up in the liquid fraction as the cheese sets (Brody, 2000). The CGMP takes up 10-20%(Brody, 2000; Daddaoua et al., 2005) of the whey protein available as a side stream product from the cheese production which is used or sold as a food additive or protein supplement(Fitzpatrick, 2012). However, due to several unique properties, the CGMP is purified and sold as a separate product. The CGMP protein sequence is rather defined with very little variation as seen in Fig.3. When inspecting the sequence it is soon realized that there is a very low level of aromatic AAs in the sequence, and thus CGMP an ideal protein source for individuals with the metabolic disorder; phenylketonuria (PKU). Numerous other beneficial effects can be attributed to CGMP as reviewed by Brody et. al.(Brody, 2000) but most importantly, with regards to HMO production, CGMP is O-glycosylated at up to five available glycosylation sites (see Fig. 1). Data was collected from literature to discuss the complete glycosylation pattern of CGMP and it suggests that five different glycosylation structures can occur(Brody, 2000). The saccharides occurring in the CGMP glycosylation pattern are anchored to the peptide by an initial GalNAc unit which can be glycosylated to form a disaccharide Gal β1→3 GalNAc. The disaccharide can be decorated with sialic acid in an α2→3 Gal and/or α2→6 GalNAc fashion, rendering a total of 5 possible glycosylation structures shown in Fig.1. It has been observed that

the sialic acid content of CGMP may vary from batch to batch and it has been shown that the variation may be due to the heat-treatment that CGMP undergoes during the industrial preparation(Li & Mine, 2004; Taylor & Woonton, 2009; Villumsen et al., 2015). However there is generally an even distribution of α2→3 and α2→6 bound sialic acid (if it has not been treated with a sialidase). As a substrate for HMO sialylation CGMP has several advantages. Whey is produced at a rate of 180 to 190×10^6 ton/year – such an amount, that large quantities are disposed of through waste channels polluting the environment and recycling of whey is therefor of interest(F, 1983). Going through the enzymatic reaction sialylating HMOs CGMP may lose some of it claimed beneficial properties, but it can subsequently serve as a protein supplement (suited perfectly for individuals with PKU) which was one of its original applications. Thus the CGMP is indeed reusable as long as process design allows for recovery of the, albeit necessary to investigated to what extend the properties of the CGMP has changed after the de-sialylation.

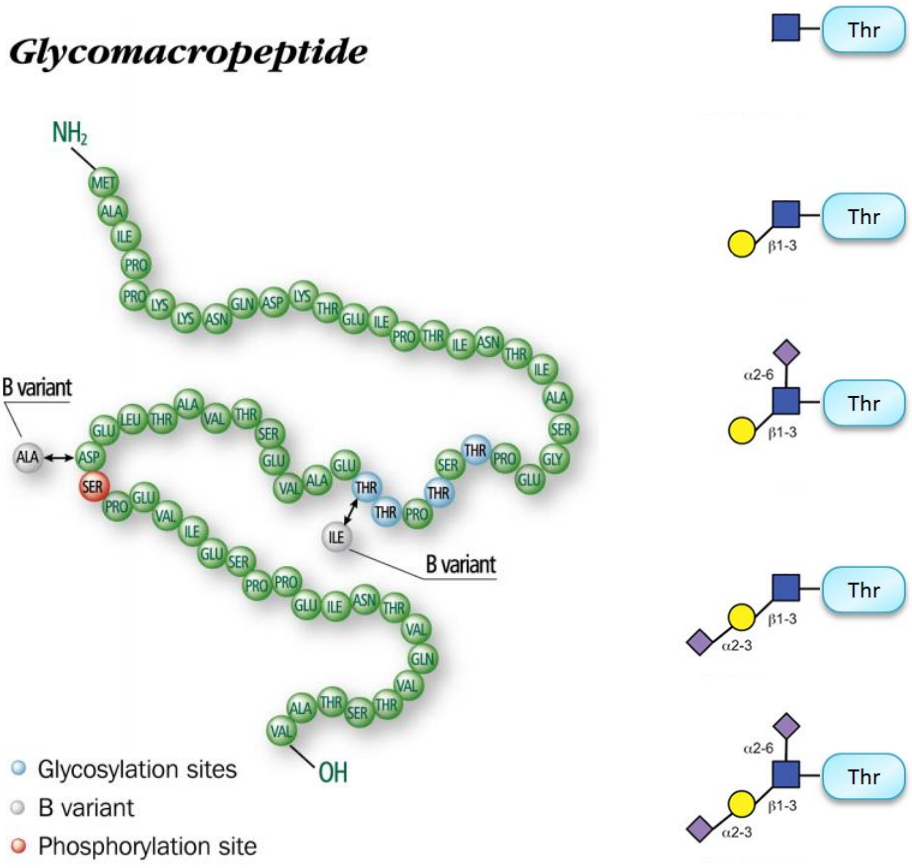


Figure 3, (adapted from www.arla.dk): The sequence of CGMP is well preserved, but occurs in two variants which differ at two positions (as indicated in grey). O-linked glycosylation can occur threonine of serine residues, but in the sequence only four of the threonine residues (blue) serve as glycosylation sites whereas the only serine serves as a phosphorylation site. Whereas there is agreement on CGMP being glycosylated at up to four sites and on the location of the first three glycosylation sites (T131, T133, and

T135) the fourth site has been reported to instead be T142 (adopting nomenclature from Eigel et al.)(Eigel et al., 1984)**. Five glycosylation structures which have been found on CGMP (right hand side) and the distribution is reportedly 0.8%, 6.3%, 18.4%, 18.5%, and 56.0%**(Saito, Yamaji, & Itoh, 1991)**.**

A fucose donor candidate – Xyloglucan

Finding a good natural source of fucose for production of fucosylated of HMOs has proven difficult and will not be discussed in the thesis beyond this paragraph. However to fully appreciate the effort that is put into the area of HMO development it is worth mentioning that a potential donor substrate, xyloglucan, has been applied for production of fucosylated oligosaccharides by BioEng (paper in preparation). Despite the successful synthesis of fucosylated oligosaccharides, the xyloglucan (mainly due to its complexity) is not an ideal substrate and the search for one continues.

Carbohydrate-acting enzymes

Since this PhD project is on the production of prebiotic oligosaccharides by novel enzymatic catalysis and since all documented prebiotics are carbohydrates(Roberfroid, 2007), this project will exclusively deal with carbohydrate acting enzymes, collectively known as CAZymes(Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014b). For the past two decades information on CAZymes have been collected and made publicly available in the CAZy database (http://www.cazy.org) (Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014). With the CAZy database followed some terminology which have been generally accepted by the scientific community. CAZymes are largely classified due to sequence similarity and sub classified in families based on significant similarity of sequence and structure(Naumoff, 2011), with at least one biochemically founding member(Henrissat, 1991). Currently the database is comprised of 6 classes of enzymes which catalyze assembly (glycosyltransferases) and degradation (glycoside hydrolase, polysaccharide lyases, carbohydrate esterases and auxiliary activities) of carbohydrates(Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014a). The sixth category (which will not be further discussed in this thesis) is carbohydrate binding modules, which occur as subunits in larger enzymes which could be hydrolases or even non-catalytic proteins(Peter Tomme, R. Antony J. Warren, Robert C. Miller, Douglas G. Kilburn, & Neil R. Gilkes, 1996) and which are mostly responsible for carbohydrate recognition(Boraston, Bolam, Gilbert, & Davies, 2004). The classification in families is generally well correlated to the catalytic mechanisms of the enzymes, but this PhD project will revolve around the exceptions to the rules – more specifically the use of enzymes from carbohydrate degrading families, the so called glycoside hydrolases (GH), to carry out carbohydrate synthesis.

Enzymes used for HMO production

Due to the increasing interest in HMO synthesis, the applicability of an array of enzymes has been addressed in recent years. Recently a thorough review of enzymes for HMO production has been published(Zeuner, Jers, et al., 2014). To produce HMOs cost efficiently, the enzymes requested for the production of HMOs must be able to utilize cheap available substrates as the ones described in the previous section. Aforementioned review therefore focuses on enzymes of different GH families, which have such structures as their natural substrates. More specifically the enzymes are retaining glycoside hydrolases, which follow the so-called classical Koshland retaining mechanism(KOSHLAND, 1953). The first step in the reaction mechanism is the formation of a covalent enzyme-substrate complex (CI)(KOSHLAND, 1953). For retaining GHs hydrolysis occurs when the CI is attacked by a water molecule facilitating the release of the hydrolyzed substrate and restoration of the active site. For trans-glycosylation to occur a hydroxyl group (e.i. from a sugar or an alcohol) must instead attack the CI(Zeuner, Jers, et al., 2014). Trans-glycosylation activity was observed for all three enzyme classes reviewed by Zeuner et al. which included were wild-type hydrolytic sialidases, fucosidases, and β-galactosidases(Zeuner, Jers, et al., 2014). However enzymes (including the ones in the aforementioned review) have more often than not been tested on e.g. para-nitrophenol (pNP) activated substrates. Thus the most common suspect enzymes; sialidases, fucosidases, and β-galactosidases have in general been tested on Sia-α-pNP, Fuc-α-pNP, and Gal-β-oNP respectively(Zeuner, Jers, et al., 2014). There are a few examples of enzymes belonging to different GH families which prefer glycoside transfer (to an acceptor)(María Fernanda Amaya et al., 2004; Montagna et al., 2002). However, hydrolysis (as the name suggests) is the prevalent activity observed within the GH families(Naumoff, 2011). The enzymes' hydrolytic activity naturally hampers their application for glycan synthesis but several examples of directing the activity of the enzymes towards reduced hydrolysis, favoring transfer-glycosylation, has been demonstrated (Jers et al., 2014; Michalak et al., 2014; Paris et al., 2005). Trans-glycosylation in the form of trans-sialylation is schematically drawn in Fig. 4. As illustrated, both donor and formed product can be subject to hydrolysis as the product is similar to the donor molecule. The trans-sialylation reaction drawn in Fig. 4 differs from other trans-glycosylation reactions (carried out by retaining GHs) only by the type of donor, (although an enzyme with the correct specificity is needed for a specific trans-glycosylation reaction).

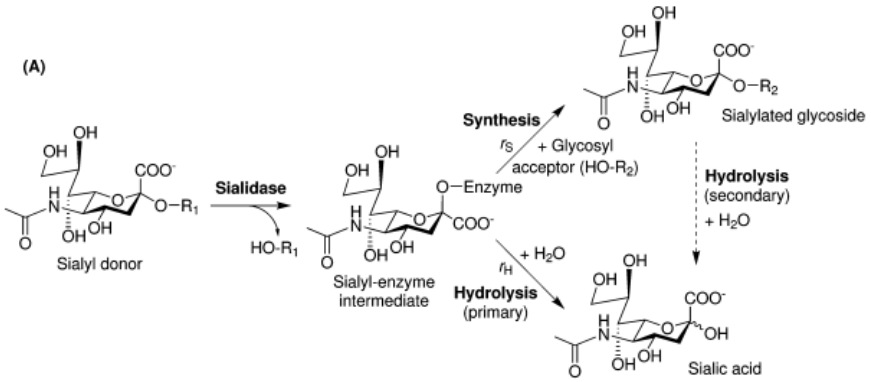


Figure 4, (Adapted from Zeuner et al. (Zeuner, Jers, et al., 2014)): In this example of a trans-glycosylation reaction by a retaining glycoside hydrolase, a sialic acid residue is first captured by the enzyme in the Sialyl-enzyme CI, before the CI is attacked by either (synthesis) a (nucleophile) hydroxyl group on a glycoside or by (hydrolysis) a water molecule. The synthesized product can also be hydrolyzed post-synthesis (Secondary hydrolysis) as indicated by the dashed arrow. For obvious reasons limiting the hydrolytic activity of GHs is a main concern when application for glycan synthesis.

Case study – TcTS

Despite having the same basic mechanistic traits as their family members, enzymes from several of the retaining GH families are capable of catalyzing the transfer of the glycoside retained in the covalent intermediate to an acceptor molecule. Some of these enzymes almost exclusively carry out the transfer reaction and one of the best studied enzyme is the trans-sialidase from *Trypanosoma cruzi* (TcTS(María Fernanda Amaya et al., 2004; Buscaglia, Campo, Frasch, & Di Noia, 2006; Holck et al., 2014; Osorio, Ríos, Gutiérrez, & González, 2012; Paris et al., 2005; Pereira, Zhang, & Gong, 1996; Smith & Eichinger, 1997)).

There are several reasons why the TcTS have received attention. Like most of the identified members of the Trypanosoma family, *T. cruzi* is a pathogen and it causes Chagas' disease in humans(Osorio et al., 2012). The TcTS has been identified as a virulence factor(Osorio et al., 2012) and several of the other pathogenic trypanosomes have similar trans-sialidases(Montagna et al., 2002; Tiralongo, Martensen, Grötzinger, Tiralongo, & Schauer, 2003). The fact that the simple mono-sialylated trans-sialylation product 3'SL and is genuine HMO may be the predominant cause of interest in this enzyme. Due to its role in pathogenesis, however, the TcTS is not "ideal" enzyme for production of sialylated HMOs. Furthermore its product specificity only allows for α2→3 sialylation(María Fernanda Amaya et al., 2004; Holck et al., 2014). The search for a trans-sialidase of non-pathogenic origin has commenced and an obvious source to investigate was the non-pathogenic trypanosome T. rangeli. However it seems to be the case that the non-pathogenic

T. rangeli, possesses only an exclusively hydrolytic sialidase (TrSA)(Pontes-de-Carvalho, Tomlinson, & Nussenzweig, 1993). The TrSA have 70% sequence similarity to the TcTS and a tertiary structure which is extremely similar(M.F Amaya, Buschiazzo, Nguyen, & Alzari, 2003), raising the question if trans-sialidases are intrinsically related to pathogenesis.

Enzyme Discovery and Development

Enzyme development is the general term for the process of acquiring novel enzymes, which can be done in a variety of refined ways. The first enzyme was discovered in 1833, Anselme Payen and Jean-François Persoz described the isolation of an amylolytic substance from germinating barley(Polaina, 2007). This discovery, however, was very different from the systematic enzyme development approaches which go on today, driven by demands from industry and societies for green, cost-efficient and sustainable solutions to increase the productivity and solve some of humanity's biggest challenges.

Many of the traditional strategies for enzyme development have been used for decades and are still in use today. However, new techniques constantly emerge not least because of the rapid progress in the field of bioinformatics where a dramatic expansion of the amount of sequenced based data drives the development of new tools – some of which are intended or suitable for enzyme development. At this point it is hard to argue that one strategy is superior and demonstrating this fact several strategies were used to acquire novel enzymes during this PhD study. All the strategies used in this study and a few more, will in the following be further discussed, highlighting their pros and cons.

At the root enzyme development can be split into two paths. One is the search in nature for enzymes which through millions of years of evolution have been developed by living organisms to serve their needs and survival. These enzymes are often referred to as wild-type enzymes which passively imply the existence of the other source of novel enzymes". Wild type enzymes are often not suitable for the harsh industrial processes and a focused enzyme development program can be necessary to achieve the desired task. Enzyme development is done by genetically modifying the coding gene-sequence to obtain an enzyme mutant with new or different properties.

In vitro enzyme discovery

The first enzymes were identified alongside or as the result of observed biological processes, and a variety of enzymes were identified in this way as science progressed in the endeavor(Bornscheuer et al., 2012). Increasingly, however, application is the main driver for enzyme identification, calling for enzymes to catalyze very well characterized reactions or solve specific problems (i.e. the transfer of a sialic acid residue

from a defined donor molecule to a defined acceptor molecule(Zeuner, Jers, et al., 2014) or to remove a stains from textiles(Niyonzima & More, 2015)).

As mentioned, enzymes are used as catalysts in *in vitro* industrial production of food, food ingredients, feed, Biofuel, Biorefinery and Biochemicals(Bornscheuer et al., 2012). Only a few of the millions of enzymes found *in vivo* in microorganisms and other higher spp. have been exploited for commercial applications. One reason is that nature's vast enzyme diversity also means that identification of "the best" enzyme (for a given reaction) is a quest, which leaves "finding the needle in the haystack" seem elementary. Hence a variety of strategies to screen for interesting enzymes have been developed.

High throughput *in vitro* screening assays

Almost all high throughput techniques for enzyme discovery rely to some extend on plate assays. The simplest plate assays rely on only the agar plate itself with addition of an organic substrate of interest. Any bacterial or fungal microorganism colony successfully growing on the plate will be capable of breaking down and utilizing the carbon source and therefore express enzymes which can break it down. This type of plate assay are referred to as a functional screening and can in a high throughput manner be used to screen libraries of bacteria or fungi capable of degrading the starch, but it can also be used to evaluate enzyme preparations.

For the plate assay to be effective in comparison of enzymes, the substrate must change visual appearance upon reaction. Such a color change can be used in two ways:

1) The substrate is colored in its native state and loses color upon enzyme action as is the case of the starch plate assay where a change from a "milky white" appearance to transparent takes place treated with an α-amylase. The clearing zone around the colony or applied enzyme sample is an indication of enzyme activity towards the breakdown of starch.
2) The substrate is colorless in its native state, but the breakdown product produces a color. An array of artificial substrate substitutes with caged chromophores have been produced for this purpose, some of which have been used in this PhD project (e.g. pNP-GlcNAc).

The artificial substrate plate assays have been used to discover a myriad of enzymes and continues to be used(Uchiyama & Miyazaki, 2009). It is however important to keep in mind, that using such assays can cause the identification of false positives and prohibit identification of false negatives, since "you get what you screen for". Furthermore the traditional use of plate assays relies on identifying enzymes in organisms which can be cultivated on agar plates. It is commonly accepted that the vast majority of microorganisms

(maybe more than 99%)(Amann, Ludwig, & Schleifer, 1995) cannot be cultivated in a laboratory setting (using the present techniques).

Metagenomic library screening assays

Different strategies are pursued to access the genes from the estimated 99% of microorganisms that cannot be cultivated. One of the most successful approaches is, however, the creation metagenomic libraries(Rondon et al., 2000). A genomic library is the entire genome of an organism transformed, as fractions inserted in identical vectors, into a monoculture of bacteria of fungi. By cloning the metagenome (which is defined as the collected genomic material from an environmental sample), in a similar fashion a metagenomic library is created(Rondon et al., 2000). The metagenomic library can subsequently be cultivated and screened using a plate assay as previously discussed and such approaches have produced enzymes of many different classes including classes of CAZymesa(Lorenz & Eck, 2005). The metagenomic library screening approach has its own drawbacks since it relies on the enzymes of interest being successfully expressed in the library host organism. Expression in the host organism depends on recognition of the native promotor preceding the foreign gene and can be effected by the lack or presence of transcription factors in the host organism(Handelsman, 2004).

In silico enzyme discovery

An alternative to the traditional wet chemistry screening assays is to search in publicly available gene sequence databases for sequences encoding relevant enzymes. Such endeavors are however simultaneously empowered and hampered by the increasing amounts of sequence data available. The amount of sequence data published in publicly available databases has since collection started increased almost exponentially(Kircher & Kelso, 2010; Lander et al., 2001; Venter et al., 2001), and continues to do so due the continuous development of new sequencing tools(Kircher & Kelso, 2010). Thus more than 16 million gene sequences are at the moment (numbers from September 2014) available as a hunting ground for in silico enzyme identification in the largest sequence database available which is the NCBI GenBank(Brown et al., 2015).

The vast amounts of generated data require automated computational tools for sequence and gene annotation. While the tools are generally effective, many sequences are labeled as hypothetical genes (e.g. ORFs which are not verifiably expressed) and other sequences are annotated incorrectly(Schnoes, Brown, Dodevski, & Babbitt, 2009). One of the more difficult features to annotate is gene product (protein) functionality(Kihara, 2011), which for obvious reasons is the key feature is relation to sequence based enzyme discovery. Functionality prediction for carbohydrate active enzymes is difficult because enzymes with low sequence identity may have similar activity(Busk & Lange, 2013), but the reverse challenge is also

encountered where enzymes with similar sequence and tertiary structure does not share substrate specificity(Yu et al., 2005). Furthermore there is the case of the trans-glycosidases: These enzymes can have almost identical active sites and tertiary structure, and share substrate specificity, while the activity is completely reversed with only minor changes to the coding sequence as it is the case for TcTS and TrSA (María Fernanda Amaya et al., 2004; Paris et al., 2005). When it comes to CAZymes, another limitation of the automated annotation is that enzymes are simply allocated to a GH groups based on a blast or HMM analysis leaving proteins without a functional classification(Bech, Busk, & Lange, 2015).

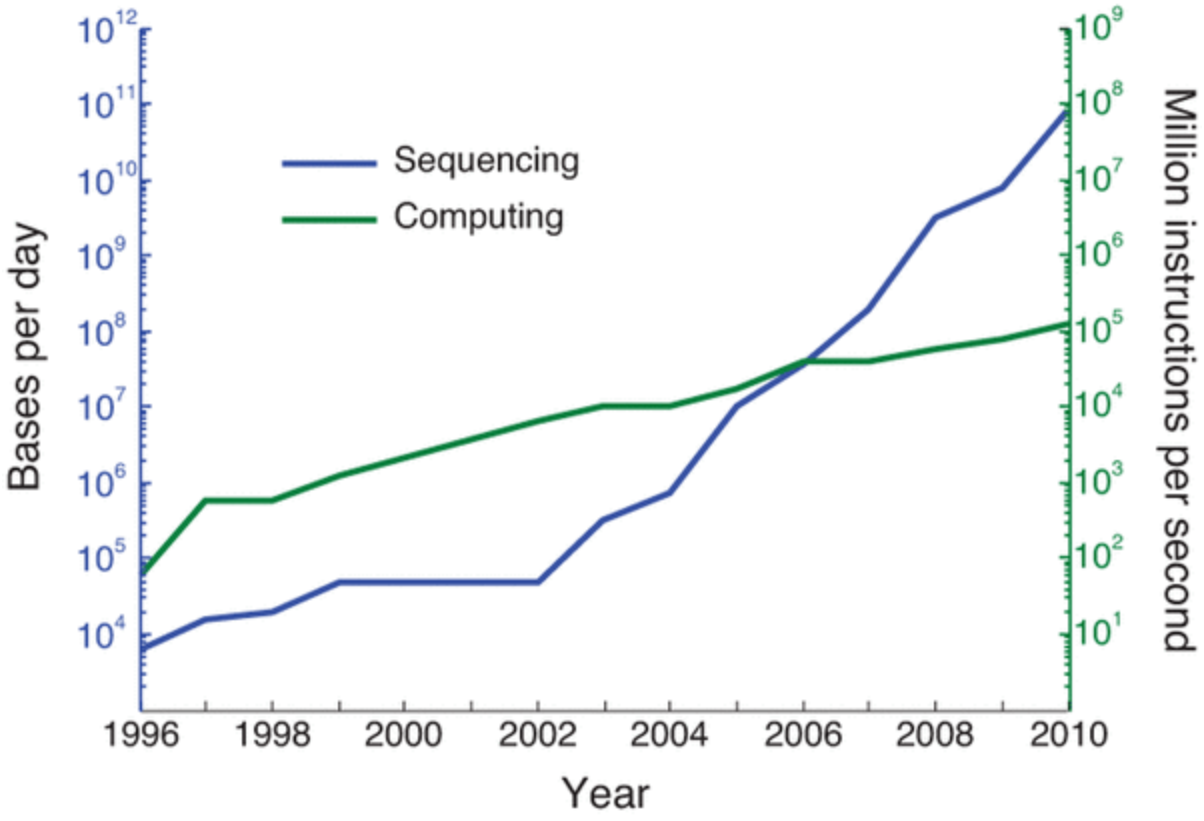


Figure 5, (Adopted from (Loh, Baym, & Berger, 2012)): In the past decades improvements in sequencing techniques have meant that sequence-data (blue, primary axis) is generated at a rate that is significantly faster than the rate in which computer power is generated (green (secondary axis). Thus computation becomes an increasingly difficult challenge and development of refined (or clever) approaches is necessary.

The enormous amount of data is not only a challenge due to error prone annotation algorithms. Searching the established databases becomes increasingly cumbersome as the data generation is outperforming the

optimization and generation of computer power(Gross, 2011; Kahn, 2011) as shown in Fig. 5(Loh et al., 2012). Thus development of sophisticated algorithms is necessary to effectively mine the sequence data for the best enzyme candidates. However the field of sequence based enzyme discovery is still in its infancy and although some interesting enzyme discovery tools have emerged(Busk & Lange, 2013; Yu et al., 2005) aiding the endeavor, BLAST analysis is still the primary tool used.

Enzyme modification by mutagenesis and directed evolution

As mentioned the alternative to novel enzyme identification is the invention of new enzymes through modification of known enzymes, optimizing the desired enzyme property. The range of properties that can be optimized is vast and there are many examples of enzyme optimization. Thus enzyme optimization have been used to successfully change the temperature optimum, change pH optimum, enhance thermostability(Miyazaki & Arnold, 1999), enhance pH stability (Koksharov & Ugarova, 2008; Kretz et al., 2004), improve substrate specificity(A, 1985; Kretz et al., 2004), enantioselectivity (Kretz et al., 2004), binding affinity (Kretz et al., 2004), protein folding(Kretz et al., 2004), protein expression(Ghazi, 2014; Kretz et al., 2004), and productivity(Kretz et al., 2004), and to infer anchoring sites for enzyme immobilization(Zeuner, Luo, et al., 2014). As it is the case for novel enzyme discovery, enzyme optimization can be done using more or less rational strategies. Rational approaches to enzyme optimization rely on insight to the structural properties of the subject enzyme and can be based on computational modeling and/or sequence investigation(Bornscheuer et al., 2012). The least rational approach to enzyme optimization is random mutagenesis. However a rational approach to random mutagenesis can be taken, utilizing the power of directed evolution, which is also widely applied to enhance the performance of rational enzyme optimization approaches. The concept of directed evolution simply covers the experimental evaluation of (a group of) obtained mutant(s) before submitting the best candidate(s) for subsequent rounds of mutagenesis, although the term precedes many of the modern techniques and have been used for experiments based on cell cultures(Francis & Hansche, 1972) experiments which may experience a revival for enzyme development(Badran & Liu, 2015).

Site saturated mutagenesis

Site saturated mutagenesis (SSM) is the experimental evaluation of the substitution on an amino acid at a specific position in a sequence by any other amino acid residue, which will naturally give rise to 20 mutants. This approach is specifically effective with regard to alteration of substrate and /or product specificity when it is used on a residue which is known to be part of the catalytic site of an enzyme. Often several sites will be tested, but it is clear that the work load is increased with the construction and handling of 19 mutants for every site that is investigated. Although SSM is a powerful tool for enzyme optimization is depends on the prior knowledge about which sites to investigate. This limitation was overcome by the invention of

Gene Site Saturation Mutagenesis™ (GSSM)(Kretz et al., 2004). In GSSM every residue in the enzyme sequence undergoes SSM without any prior knowledge to the structure or mechanism of the enzyme(Kretz et al., 2004). Contrary to SSM(Kretz et al., 2004), GSSM can efficiently be used in a directed evolution set-up, but is nonetheless relying on the effect of the inferred mutations being somewhat additive.

Random mutagenesis

All enzyme optimization rely on the alteration of the protein sequence of the enzyme. Traditionally in random mutagenesis the DNA material coding for the enzyme is damaged by a chemical agent(Myers, Lerman, & Maniatis, 1985; Ohnuma et al., 1996) or radiation(Ghazi, 2014). The mutagenesis is completed when the expression host organism tries to repair the DNA, but fails to incorporate the correct base in the sequence. Random mutagenesis have also been carried out by the use of an error-prone DNA-polymerase, introducing in random fashion during PCR amplification of an enzyme gene(Cadwell & Joyce, 1992). The properties of the mutated enzymes will most likely be worse than the wildtype and it is therefore necessary to identify improved enzyme mutants using one of the aforementioned assays. For decades random mutagenesis have been used to obtain enzymes with enhanced properties d(Autenrieth & Ghosh, 2015; M, 1988; Moore & Arnold, 1996; Zaccolo & Gherardi, 1999), although the strategies for obtaining the mutants have developed throughout. As have been, and will be described, an array of refined methods for enzyme modification has been developed. However there are several advantages to random mutagenesis meriting its continued use - for example little or no knowledge about the assayed enzyme is needed. Furthermore it is possible to identify improvements caused by complex non-additive poly-mutated specimens, which might be lost in approaches where individual mutations are evaluated.

Sequence Alignment and Analysis

Sequence analysis is the corner stone in almost all rational design and is encompassed in the enormous field of Bioinformatics. As mentioned in relation to *in silico* screening the access to sequence data is ever increasing, but when it comes to using sequence analysis for enzyme optimization only sequences with descriptive information is of use. In essence parallels are drawn between the enzyme of interest and any sequence or motif described in literature.

As discussed BLAST can be a very useful tool for enzyme identification, but it can also assist in the development of rational enzyme optimization strategies. Based on BLAST analysis conserved residues, which due to their conservation may be important to functionality, can be identified. Naturally such analysis works better if a collection of related sequences are known. Based on the similarity, strategies for enzyme optimization can be developed as exemplified in the case study on Tr13. Furthermore, a BLAST analysis can be used to identify which purposes individual domains serve (for multi-domain enzymes).

3D structure analysis and homology modeling

Another tool that has become increasingly popular for protein optimization is the analysis of the 3-dimentional structure of proteins. An accurate 3D structure of a protein is determined by either X-ray or NMR analysis of the enzyme (Krishnan, V and Rupp, 2012). X-ray analysis relies on a successful crystalline preparation of the enzyme of interest, which in some cases can be difficult to obtain(Krishnan, V and Rupp, 2012). However the X-ray analysis will in turn generally produce models of the 3-dimentional structure of the protein of better resolution(Krishnan, V and Rupp, 2012). Whereas great amounts of work is required to obtain 3-dimentional structure models experimentally existing protein models may serve as scaffolds to obtain so called homology models(Nielsen, Lundegaard, Lund, & Petersen, 2010; Söding, Biegert, & Lupas, 2005). The quality of the models created can vary with the homology to, and quality of the reference model (especially in the detail), but the overall tertiary structures are generally predicted well since even remotely homologous enzymes share 3-dimensional structure(Kinch & Grishin, 2002).

The visualization of an enzymes 3-dimentional structure compliments many of the aforementioned tools and techniques, but is a brilliant tool on its own as well. Whereas the secondary structure of a protein is difficult to comprehend and difficult to intuitively translate to functionality 3-dimetional models are ideal for this purpose. Thus the visual inspection of the 3-dimensional structure of the TcTS was the prime inspiration to the study in Chapter two, where also homology modeling played an important role.

Computational modeling

Both sequence and 3-dimentional structure analysis can be the basis for an ever increasing array of computational tools for protein optimization(Bornscheuer et al., 2012). To highlight an example the PoPMuSiC algorithm(Gilis, 2000) has repeatedly demonstrated its use for increasing the stability of proteins(Zhang & Wu, 2011). However increased stability is not necessarily the same as increased activity, the loss of which may sometimes be the cost of the increased stability, when applied on enzymes. Computational modeling can be extremely refined as exemplified in a study by Pierdominici-Sotile et al.(Pierdominici-Sottile, Palma, & Roitberg, 2014) which was the basis for the study described in Chapter 4.

Case study – Tr13

An example of BLAST analysis applied for protein optimization is the further development of the engineered trans-sialidase, Tr6, which was engineered from the exclusively hydrolytic sialidase of *Trypanosoma rangeli* (Paris et al., 2005). Based on a BLAST analysis where this sialidase was aligned to trans-sialidases found in its close cousins; *T. cruzi*, *T. brucei* and *T. congolense*, an enzyme optimization strategy was developed. In this case the sequence target that was identified was a stretch of 7 AA, where the Tr6 differed significantly from its cousins (Fig. 6). By introduction of the 7 AA stretch into the Tr6 the

hydrolytic activity of the enzyme was dramatically reduced and the superior Tr13 was created(Jers et al., 2014).

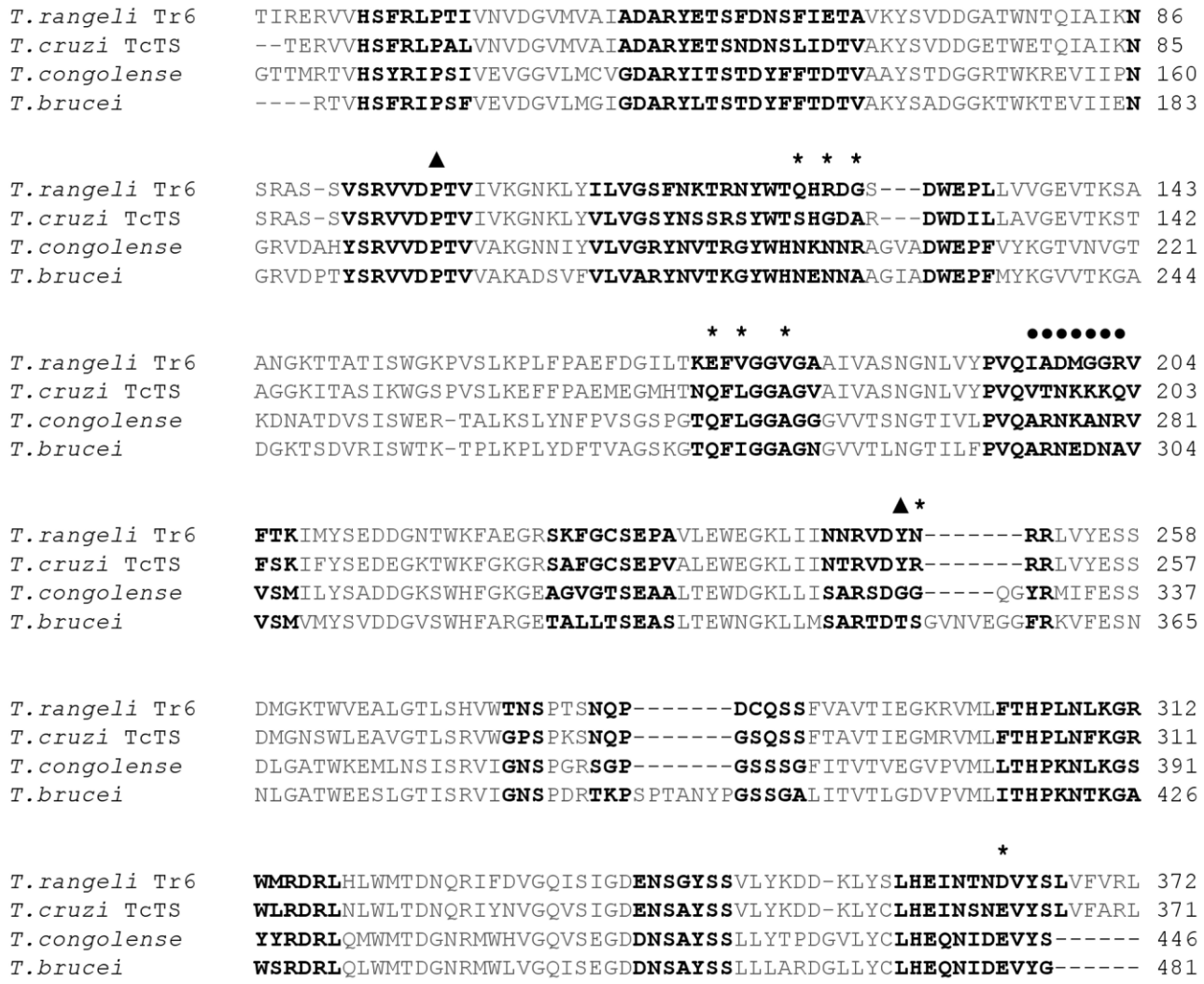


Figure 6, (adopted from Jers et al.(Jers et al., 2014)): The above multiple alignment resulted in the identification of a seven AA residue stretch (indicated by ●) where the hydrolytic sialidase TrSA differ from the efficient trans-sialidases of *T. cruzi*, *T. congolense* and *T. Brucei*. Upon genetic modification of an already genetically modified TrSA mutant (with trans-sialidase activity) the hydrolytic activity was drastically decreased rendering an superior trans-sialidase.

Processes development in relation to HMO production

In an effort to follow the HMO production (3'SL production) through to completion, the process was considered as a whole. Since the product is a high bulk product process optimization is paramount and since it is a food grade product sanitary and food grade processes are necessary. The research studies that were borne as a result of these reflections are unraveled in Chapters 5 and 6. Where Chapter 5 considers

product purification, Chapter 6 considers integrating the entire HMO production and purification process into one. However the proposed solutions to the identified challenges in both Chapters include membrane separation technology.

Ultra- and Nanofiltration

Both ultra- and nanofiltration (UF and NF) are commonly carried out as pressure driven membrane filtrations. Pressure is applied across a porous membrane resulting in solvent and solutes smaller than a specific cut off (determined by the membrane and the setup) pass through the membrane. Thus two fraction (permeate and retentate) can be recovered. Ideally the permeate fraction is without impurities (depending on the stringency of the cutoff and the composition of the sample). This is not the case for the retentate fraction which for completely un-retained solutes contains the same concentration as in the feed solution whereas retained solutes are up-concentrated. The difference between UF and NF is primarily commonly thought of as a difference in MW cutoffs of the membranes. However there is an overlap in the cutoffs as the difference between UF and NF is actually more accurately a difference in separation behavior. As mentioned several factors can affect the molecular weight cutoff (MWCO)size of a membrane. For UF separation is almost exclusively depending on particle and pore size, whereas other factors such as osmotic pressure, hydrophobicity and electrostatic interaction play important roles in NF.

Biocatalytic productivity

For the optimization of enzymatic reactions one of the best measures to evaluate is the biocatalytic productivity which is the amount of produced product per amount of enzyme (on a weight basis unless stated otherwise). Optimization of the biocatalytic productivity aims to reduce the cost of enzyme, which is often a large proportion of the total cost of an enzymatically driven process. The biocatalytic productivity is a valuable measure since it encompasses the process as a whole on an enzyme basis. However whereas the measure makes comparison between processes easy it is important to appreciate any differences in starting and end point. Furthermore it is immediately understood that other measures must be evaluated to compare processes, such as energy and water consumption, time, manpower and factory volume to mention a few. For example heavy pretreatment may improve biocatalytic productivity drastically, whereas the adhered energy consumption may render such pretreatment completely unmerited.

Improvement of biocatalytic productivity is however not always accompanied by disadvantages. By the use of enzyme membrane reactors, where product is continuously removed have been shown to be an effective way of improving biocatalytic activity in reactions where the formed product inhibit the reaction(Gavlighi, Meyer, & Mikkelsen, 2013) or where the product is not stabile in the reactor. Since the central theme in this thesis, enzymatic synthesis of prebiotic saccharides by GH family enzymes, involve an

imminent risk of product hydrolysis taking place, enzymatic membrane reactors are useful in optimizing biocatalytic productivity as it has been shown by Zeuner et al. for 3'SL production(Zeuner, Luo, et al., 2014). Biocatalytic productivity may also be increased by immobilization, which apart from allowing enzyme recycling, also may increase enzyme stability and/or activity – trans-sialidase-relevant examples include immobilization on concanavalin A-sepharose(Scudder, Doom, Chuenkova, Manger, & Pereira, 1993; Shiian et al.), VA-epoxy(THIEM & SAUERBREI, 2010) and on Cu^{2+}-iminodiacetic acid (IDA)-functionalized carbon-coated magnetic nanoparticles (MNPs) via His-tag, which enable enzyme recollection by magnetic attraction(Zeuner, Luo, et al., 2014).

Hypotheses

Hypothesis 2.1: The family of sialidases is extremely conserved and it can be speculated that trans-sialidases could have developed in parallel with the TcTS inferring aromatic sandwiches in similar positions as it is found in TcTS.

Hypothesis 2.2: If hypothesis 2.1 is correct, sialidases with an aromatic sandwich above the active site will be excellent trans-sialidase candidates.

Hypothesis 2.3: A novel identified trans-sialidase will be able to produce 3'SL from CGMP and lactose in the same way as TcTS.

Hypothesis 3.1: Because GlcNAc in the polymeric form of chitin as exoskeleton in many lower eukaryotes such as insects and arachnids inhabiting soil, occurrence of GHs capable of degrading GlcNAc structures are expected to be possible to identify from a soil derived metagenomics library.

Hypothesis 3.2: Among GH20 enzymes identified from metagenomics libraries, specimens can be found which are able to transfer a GlcNAc residue to various acceptor molecules.

Hypothesis 4.1: The mutations suggested by Pierdominici-Sottile et al.(Pierdominici-Sottile et al., 2014) will improve trans-sialidase activity in TrSA mutants, but the trans-sialidase activity will be further improved by combination with the loop mutations suggested by Jers et al.(Jers et al., 2014).

Hypothesis 4.2: The improved trans-sialidase activity will enable efficient trans-sialylation at reaction conditions with low levels of acceptor substrate.

Hypothesis 4.3: By implementation novel mutations in groups it will be possible to gain knowledge about the mechanistic traits on the individual mutations (in groups) contribution to trans-sialidase activity.

Hypothesis 5.1: Since NF efficiency can be heavily affected by the membrane zeta potential (charge), and since the charge difference between lactose and 3'SL is large, NF it will be an efficient tool for separating 3'SL and lactose despite similar size of the molecules (from a high molecular weight NF perspective).

Hypothesis 1: The following observations from large scale cross-flow UF unit will translate well to a laboratory experimental setup using dead-end stirred tank filtration setup:

- Fouling is the main concern with regard to CGMP UF.
- Hydrophobic membranes are unsuitable for CGMP filtration due to heavy fouling.
- UF is a viable method for separation of product mixture (3'SL and lactose) from the remaining reaction components.

Hypothesis 2: An additional hypothesis that was tested in this study was that the biocatalytic productivity could be greatly enhanced by application of the integrated membrane system as enzyme could be retained in the EMR and reused accordingly.

Chapter 2: A Rational Approach to Identification of Wild Type Sialidases with "Trans-sialidase" Activity for the Production of Human Milk Oligo Saccharides

2.1 Motivation

Throughout this PhD study, the enzymatic reaction which has received most attention is the sialylation of lactose. Most often genetically modified specimens of the TrSA have been used, but while progress has been made these mutants still rely on high acceptor concentrations in the reactions to eliminate hydrolysis(Jers et al., 2014; Michalak et al., 2014). As mentioned the most efficient trans-sialidases are the trans-sialidases found in the trypanosomal species with the TcTS being the most studied. However the use of the TcTS for industrial purposes is problematic for several reasons: 1) The TcTS has been identified as a virulence factor and extensive purification and quality control would be necessary before any component of a reaction using this enzyme could be used (or reused in the case of the starting materials). 2) In our research group expression of the TcTS has proven difficult and much work will go into enzyme production at decent levels. 3) The use of the TcTS for HMO production is protected by patents. 4) TcTS cannot synthesize 6'-sialylated compounds.

The primary motivation for this study was the challenge of identification of a genuine trans-sialidase, which did not originate from the Trypanosomas genus. Through reading this PhD thesis it is clear that focus on the challenge is rooted in the application for HMO production. However, the following paper is written with a different scope. The scope was changed because of the approach, taken to develop the screening strategy, could have a broader impact and hopefully inspire a wider audience within the field of enzyme discovery (more specifically scientist working with identification of trans-glycosidases).

2.2 Hypotheses and objectives

The research leading to the following paper had a very clear objective; to identify a novel, non-trypanosomal trans-sialidase. The paper describes very well how this objective was approached, and how the approach resulted in construction of several hypotheses.

Hypothesis 2.1: The family of sialidases is extremely conserved and it can be speculated that trans-sialidases could have developed in parallel with the TcTS inferring aromatic sandwiches in similar positions as it is found in TcTS.

It is clear that hypothesis 2.1 does not perform well as a scientific hypothesis as it is very hard to disprove, but if it is true another more experimentally approachable hypothesis can be strung:

Hypothesis 2.2: If hypothesis 2.1 is correct, sialidases with an aromatic sandwich above the active site will be excellent trans-sialidase candidates.

Hypothesis 2.3: A novel identified trans-sialidase will be able to produce 3'SL from CGMP and lactose in the same way as TcTS.

2.3 Experimental considerations

In this study a structure based approach to identification of new trans-sialidases was rapidly designed. The screening for enzyme candidates was done in silico and the first step was carried out on sequence basis identifying a search motif. The role of the search motif was to identify whether an aromatic residue was present at a specific position in a specific loop in each enzyme. Undoubtedly a more refined strategy, first identifying the region, then the loop and eventually the presence of the aromatic residue could have been developed using delicate bioinformatics tools, but one of the messages of the following paper is the essentiality of quick verification of predictions using wet chemistry thereby justifying further endeavors to refine predictions. Endless time could have been spend developing advanced screening tools but instead It was determined that a premise of the study would be strategic focus on simplicity and prove of concept.

With simplicity in mind, the development of the search motif was more iterative and less stringent than it may appear in the paper. Search motifs were tried, changed and refined until a suitable enzyme candidate was identified. However, it is the hope that other members of the scientific community will be inspired by the intrinsic simplicity of the applied method and apply similar strategies to identify other trans-sialidases or trans-glycosidases in general. Additional enzymes are needed to build a foundation enabling development of more refined methods that could lead to a breakthrough in the field of trans-glycosidase discovery and consequently in the field of glycobiology.

2.4 Conclusions

The following paper will report on the identification of the first identified non-tryponosomal trans-sialidase and thus the objective of the study was fulfilled. The enzyme was identified based on a rational in silico screening strategy tailored specifically to identify such and enzyme. Whereas the strategy is not believed to be widely applicable, its simplicity and efficiency should be an inspiration to develop similar simple, tailored approaches. The aromatic sandwich above the TcTS active site proved successful as a trans-sialidase marker and more such structures should be screened for with the aim of identification of other trans-sialidases or trans-glycosidases in general.

The identified enzyme was capable of producing 3’SL from CGMP and lactose, but in In the process of evaluating the reaction performance a novel trans-sialylation product, 3SL, was identified. This product has not previously been described and its properties are not known.

2.5 Paper I (manuscript)

It all starts with a sandwich

A rational 3D alignment approach to the identification of sialidases with trans-glycosylation activity

Nordvang RT, Nyffenegger C, Holck J, Jers C, Zeuner B, Sundekilde UK, Meyer AS, Mikkelsen JD

Introduction:

In the database for carbohydrate acting enzymes, CAZy, more than 265000 identified glycosyl hydrolases (GH) have been grouped into families (http://www.cazy.org/Glycoside-Hydrolases). GH enzymes are allocated to a family in the CAZy database primarily based on sequence similarity, however the catalytic mechanism of enzymes within a family also tend to be conserved(Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014). A group of GH families, known as retaining glycosyl hydrolases, cleave a glycosyl residue from a saccharide through the formation of a covalent intermediate (CI)(KOSHLAND, 1953). For hydrolysis to occur, the CI must be attacked by a water molecule(KOSHLAND, 1953). However, a recent trend in the field of glycobiology, is the screening for trans-glycosylation activity among enzymes from the glycosyl hydrolase (GH) families(Zeuner, Jers, Mikkelsen, & Meyer, 2014). In trans-glycosylation, the glycosyl residue retained in the CI is attacked by a hydroxyl group of an acceptor molecule (e.g. a sugar or alcohol) through which transfer of the glycosyl residue occurs. Thus a trans-glycosidase can be defined as a glycosidase whose primary activity at a given set of conditions is the transfer of a glycoside residue from a donor molecule to an acceptor molecule.

Successful identification of trans-glycosidases will enable efficient tailoring of a myriad of refined glycosides(Bissaro, Monsan, Fauré, & O'Donohue, 2015; Rather & Mishra, 2013). In nature, saccharide-synthesis is typically carried out in a bottom-up fashion by glycosyltransferases. However, glycosyltransferases are not suitable for industrial application as they require the use of activated glycosides as substrates(Desmet et al., 2012). In the trans-glycosylation reaction, however, a specific monosaccharide is transferred from one glycan (the donor) to another glycan (the acceptor). Therefore,

trans-glycosidases are preferable to glycosyltransferases for large scale production of glycosides because they can utilize non-nucleotide-activated substrates. Furthermore, trans-glycosidases offer a higher product to waste ratio compared to traditional glycoside hydrolases (Rather & Mishra, 2013).

A primary benefit of using trans-glycosidases compared to glycosyltransferases for the synthesis of glycosidases is the ability of trans-glycosidases to utilize various, cheap forms of biomass (i.e. from industry side streams) as feedstocks. Development of trans-glycosylation processes that sequentially extract specific components from biomass, without destroying the remainder of the biomass, would enable the production of high value products from low-value biomass side streams before designating the remainder of the biomass for the production of lower value commodities such as bioethanol.

Despite the high interest in trans-glycosidases, so far their rate of identification has been slow, mainly due to complications in the screening process. Experimentally, GH enzymes are often identified using chromogenic test substrates amenable to high throughput screening, in which a chromophore is released irrespective of whether the monosaccharide is transferred to water (hydrolysis) or to a glycan (trans-glycosylation), therefore preventing specific identification of trans-glycosylation enzymes. Very few screening assays for the specific identification of trans-glycosylation enzymes have been suggested and consequently the trans-glycosidases identified thus far have generally been identified via different routes. Therefore, it seems enticing to use bioinformatics to search for trans-glycosidases, but this effort is hampered by the very limited number of characterized trans-glycosidases in different GH families and the fact that trans-glycosidases tend to be structurally more closely related to hydrolytic enzymes of the same GH family than to trans-glycosidases of different families (Bissaro et al., 2015).

Investigating the mechanisms of enzymes from the families of retaining GHs (which have been shown to possess trans-glycosylation activity), it is clear that the formation of a covalent glycoside-enzyme intermediate (CI) is the mechanistic trait enabling trans-glycosylation activity. As mentioned, hydrolysis occurs when a water molecule attacks the CI, whereas transfer occurs when it is attacked by a hydroxyl

group from another glycoside. Among the questions to be answered are therefore: What differentiates a trans-glycosidase from its hydrolytic relatives? Since trans-glycosidases are highly related to the hydrolytic enzymes within the GH family in which they occur – more so than to trans-glycosidases of other GH families – it seems convincing that subtle molecular adjustments rather than major modifications lead to the evolution of trans-glycosidases(Bissaro et al., 2015). It is therefore questionable if any traits are conserved among trans-glycosidases, but in this study we attempted to identify structural traits which can be used to identify new trans-glycosidases.

This study aimed to identify a trans-glycosidase activity marker through the investigation of a specific trans-glycosidase, namely the trans-sialidase of the protozoan *Trypanosoma cruzi* (TcTS). *T. cruzi* is a human pathogen and the TcTS has been identified as a virulence factor(Schenkman, Vandekerckhove, & Schenkman, 1993). The TcTS contributes to the pathogenicity of the *T. cruzi* by several mechanisms (Osorio, Ríos, Gutiérrez, & González, 2012), but remarkably it facilitates the coating of the protozoa mucins with sialic acid residues originating from glycosylation structures on the human epithelium, in an effort to escape the human immune system (Buscaglia, Campo, Frasch, & Di Noia, 2006). The 3D structure and mechanism of the TcTS was first reported in 2002 by Buschiazzo et al.(Alejandro Buschiazzo, Amaya, Cremona, Frasch, & Alzari, 2002) and to date the TcTS is the best characterized trans-sialidase. More trans-sialidases have been identified, but so far they all originate from the *trypanosoma* genus (Amaya et al., 2004; Ammar, Plazolles, Baltz, & Coustou, 2013; Montagna et al., 2002; Tiralongo, Martensen, Grötzinger, Tiralongo, & Schauer, 2003).

The TcTS constitutes an ideal case study for several reasons. Firstly, it is extremely selective towards trans-sialylation, with no hydrolytic activity. Furthermore, the completely hydrolytic sialidase of *T. rangeli* (TrSA), which is70 % homologous with the TcTS, has recently been engineered to exclusively carry out trans-sialylation under specified reaction conditions. Trans-sialidases also have commercial interest as it has been demonstrated that both TcTS and mutants of the TrSA can efficiently sialylate a variety of glycans, enabling

the production of sialyllated human milk oligosaccharides (Holck et al., 2014; Jers et al., 2014; Michalak et al., 2014; Paris et al., 2005) – a group of molecules that serves a multitude beneficial effects in human breastmilk (Bode, 2012) and which are desired as a supplement to infant formula (Bode, 2012).

In the TcTS, two residues above the active site, forming an aromatic sandwich, have been reported to contribute to the trans-sialyllation activity. A TcTS mutant without one half of the sandwich (Tyr119) was created and less than 1% trans-sialidase activity was maintained (A Buschiazzo et al., 2000). Furthermore, a corresponding tyrosine residue is missing in TrSA and has been introduced into all mutants of this enzyme successfully displaying trans-sialidase activity(Jers et al., 2014; Michalak et al., 2014; Paris et al., 2005). Through investigation of the TcTS structure model, the aromatic sandwich was identified as a suitable trans-sialidase identification marker, due to its specific structural characteristics facilitating the definition of a search motif. The hypothesis that was tested in this study is therefore: A sialidase with an aromatic sandwich above the active site, as it is found in TcTS, will likely be a trans-sialidase. In an iterative, *in silico* screening processes, a database of 2909 sialidases was gradually reduced to four candidates, one with an aromatic sandwich and three with similar (narrow) binding clefts, but missing one of the aromatic sandwich residues. Recombinant expression of the candidates followed by enzymatic activity studies were used to evaluate the identified candidates. The candidate displaying an aromatic sandwich, a sialidase from *Haemophilus parasuis*, indeed proved to be a trans-sialidase.

Material and methods:

In Silico Screening

The initial library dataset was assembled by acquisition of all genes in the NCBI databank matching a description of 'sialidase' or 'neuraminidase' (putative enzymes were included) and redundant sequences were removed.

Trans-sialidase search motif

As described in the results and discussion section, a search motif was created to identify possible trans-sialidase candidates. Adopting the nomenclature proposed by Aasland et al.(Aasland et al., 2002), the 5 amino acid search motif is given as ΩxRDR, where 'Ω' represents an aromatic amino acid and 'x' is any residue. Thus, the search motif identified the TcTS with W312 and Arg314 (bold) as follows:

Motife:	Ω x R D R
	● ● ● ●
TcTS (W312 - R316):	**W** L **R** D R

The terminal residues ('DR') were included in the very short search motif to achieve sufficient specificity and enable a meaningful search.

Crystal structure based protein models

Two crystal structure based protein models were downloaded from the RCSB PDB website (http://www.rcsb.org)(Berman, 2000). Namely, a TcTS model (PDB ID: 1sOI)(Amaya et al., 2004) and a model of the *Micromonospora viridifaciens* Neuraminidase (MvSA) (PDB ID: 1W8O)(Watson, Newstead, Dookhun, Taylor, & Bennet, 2004).

Modeling & 3D Alignment

Candidate enzymes were modeled by the HHPred server(Söding, Biegert, & Lupas, 2005) using the TcTS as the template when possible (otherwise, the top 5 template hits were used). Subsequently, each individual model was aligned to the TcTS crystal structure model using the alignment function of the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

Enzyme production

The candidate genes as they appear in genbank were prepared for expression in the following manner: Signaling sequences were identified and removed. In front of the protein coding sequence (at the 3'- end of the sequence) a His-tag sequence, followed by a protease site sequence were added. The resulting genes were synthesized and inserted in the pJ414 expression vector allowing for IPTG induction upon introduction into electro-competent *E. coli* BL21 DE3 pLysS by DNA2.0. Transformant colonies were picked, and used to inoculate 1 l of auto-induction medium ZYM-5052(Studier, 2005) containing 100 µg/ml ampicillin for 16 h with shaking at 25 °C in a 5 l shake-flask. The cells were harvested by centrifugation (20 min at 5000 g) and re-suspended in binding buffer (20 mM citrate-phosphate buffer, 100 mM NaCl, 15 mM imidazole, pH 7.4). Cells were lysed by addition of lysozyme, followed by a10 min period of incubation on ice and 6 X 30 s sonication, separated by equal breaks. The cell debris were pelleted by centrifugation at 5000 g for 20 min. The supernatant was subjected to sterile filtration through a 0.45 µm filter and subsequently loaded onto a 1 mL Ni^{2+}-sepharose gravitation column and washed by addition of 10 ml of a wash buffer containing 24 mM imidazole, 100 mM Nacl and 20 mM citrate-phosphate buffer (pH 7.4). The protein was eluted from the column in 3 ml of elution buffer containing 500 mM imidazole, 100 mM NaCl and 20 mM citrate-phosphate buffer (pH 7.4). Imidazole was removed from the samples by the means of GE Healthcare PD10 columns using the manufacturer's instructions, although only 3 ml of the 3.5 ml final elution volume was collected to ensure no imidazole in the samples. The purity of the final enzyme preparations were confirmedby SDS-PAGE (not shown).

Time course experiment

CGMP was purified as described by Luo et al. (Luo et al., 2014). Lactose was added to the CGMP solution to obtain final concentrations of 137.5 g/L and 50.6 g/L, respectively. Enzyme preparations were diluted to equal concentrations of total protein and an appropriate volume was heat inactivated at 90 °C for 30 min.

For each time point (see Fig. 4), reaction duplicates with active enzyme were started by addition of 135 µl enzyme to 365 µl substrate mix (both preheated to 30 °C) and the resulting reaction mixture (100 g/L lactose, 37 g/L CGMP corresponding to approximately 4 mM 3'-bound and 4 mM 6'-bound SA) was incubated at 30 °C with shaking (700 rpm). Inactivated enzymes were used as controls. Reactions were stopped by heat inactivation at 90°C for 10 min before the reaction was transferred to a 5KDa viva spin filter (Sartorius) and centrifuged for 10 min at 5000 g and 4 °C. As heat inactivation was unsuccessful for the SialA enzyme, these samples were stopped directly by 5Kda filtration, as described above. To evaluate the trans-sialylation activity of the enzymes, product to hydrolysis ratios were calculated as:

$$R_{P,H} = \frac{C_P}{C_{SA}}$$

Where (at a given time point) C_P is the concentration of a given SL variant and C_{SA} is the concentration of SA. A trans-sialidase was defined as an enzyme exhibiting a $R_{P,H} > 1$ at any time point.

Sample analysis by HPAEC-PAD

Concentrations of 3'SL, 6'SL and SA in the reaction permeate were measured by high-performance anion exchange chromatography with pulsed aerometric detection (HPAEC-PAD) analysis using a CarboPac™ PA100 (4 mm × 250 mm) analytical column equipped with a CarboPac™ PA100 (4 mm × 50 mm) guard column (Dionex Corp., Sunnyvale, CA) on a Dionex ICS-3000 system (Dionex Corp., Sunnyvale, CA). The operating conditions and analysis procedure have been previously described by Zeuner *et al.*(Zeuner, Luo, et al., 2014), and resulted in base line separation of the analytes. SA (Sigma Aldrich), 3'SL (CarboSynth) and 6'SL (CarboSynth) were used to prepare HPLC standards. For novel reaction products where no standard was available, the product formation was calculated using the extinction coefficient of 3'SL as a best estimate.

Anion exchange chromatography

Anion exchange chromatography for the purification of sialylated compounds prior to MALDI-TOF and NMR analysis was performed with an ÄKTA purifier 100 equipped with a SepharoseQ column (GE Healthcare Biosciences Uppsala, Sweden), as described previously (Holck et al 2014).

Matrix-assisted laser desorption/ionization - time of flight (MALDI-TOF)

A 1 µL sample and 1 µL of 10 mg/mL norharmane matrix in 50% acetonitrile was applied to a MTP 384 target plate ground steel BC (Bruker Daltonics), mixed on the target and dried under a stream of air. The samples were analyzed with an Ultraflex MALDI-TOF/TOF instrument (Bruker Daltonics GmbH, Bremen, Germany). The instrument was operated in negative acquisition mode and controlled by the FlexControl 3.3 software package. All spectra were obtained using the linear mode with an acceleration voltage of 20 kV, and pulsed ion extraction of 10 ns. The acquisition range used was from m/z 400 to 2500. Post-source decay (PSD) spectra using the Bruker Daltonics LIFT system were recorded at 8 kV precursor ion acceleration voltage and fragment acceleration (LIFT voltage 19 kV). The reflector voltage 1 and 2 were set to 29.6 and 13.9 kV, respectively.

NMR spectroscopy

One-dimensional ^{1}H, 2D $^{1}H^{1}H$ and $^{1}H^{13}C$ NMR spectra were acquired at 298 K on a Bruker Avance III 600 spectrometer, operating at a ^{1}H frequency of 600.13 MHz, equipped with a 5-mm 1H TXI probe (Bruker BioSpin, Germany). Proton and carbon chemical shifts were measured in reference to an internal acetone standard (Sigma-Aldrich; ^{1}H δ 2.225 and ^{13}C δ 31.08 ppm). After synthesis samples were lyophilized and redissolved in 650 µL D_2O (Sigma-Aldrich). 1D ^{1}H NMR spectra were recorded with a spectral width of 6000 Hz, and a total of 64 scans were collected into 32 K complex data points. Water signal was irradiated using pre-saturation. $^{1}H^{1}H$ COSY spectra were recorded with 256 increments in 4096 complex data points with a spectral width of 6000 Hz. $^{1}H^{13}C$ HSQC spectra were recorded with a spectral sweep width of 6000 Hz in t_2

and 22637 Hz in t_1 direction. All spectra were acquired and processed using Topspin 3.0 (Bruker Biospin). With regard to evaluation of spectra, all resonances are referenced to internal acetone (δ 2.225).

Results & Discussion

TcTS analysis and Trans-sialidase identification strategy

A visual representation of the TcTS Michaelis complex (MC) with 3'SL is presented in Figure 1 and an overview of residues interacting with the substrate in the MC is given in Table 1. The catalytic residues and the residues responsible for sialic acid moity fization are conserved among all sialidases but as mentioned previously, the primary identification marker chosen for the screening strategy was the aromatic sandwich comprised of W312 and Y119. Several factors pointed to the aromatic sandwich as a good trans-sialidase identification marker. As mentioned previously, the aromatic sandwich has been reported to be essential to the trans-sialidase activity of the TcTS. However, it is the relative distance of the aromatic sandwich to another specific residue, R314, which renders the aromatic sandwich a suitable selection marker. Whereas the two residues comprising the aromatic sandwich are unique to trypanosomal trans-sialidases, the active site residues in Table 1 are all highly conserved among all sialidases. Therefore R314 is of special significance with respect to the screening strategy, since it is positioned only two residues from W312 of the aromatic sandwich. The close proximity of the two residues enabled the creation of a search motif which could be used for a database reduction.

The initial reduced database ideally identified all enzymes with an aromatic residue in a position (relative to the arginine residue) where an aromatic sandwich could be established if a corresponding aromatic amino acid is present in the enzyme. Thus the remaining part of the screening strategy was focused on determining the presence of "the other half" of the sandwich and was initiated by the construction of homology models of the sequences in the remaining sequences. Alignment of the constructed models to

the TcTS and visual evaluation was the final step of the screening process where candidates were selected based on two modelling-criteria: 1) restoration of the active site with its 9 conserved residues (see Figure 1) and 2) the formation of a narrow (shielded) binding cleft above the active site, preferably with an aromatic sandwich composed of two aromatic amino acid residues.

Evaluation of initial database reduction

The initial database reduction resulted in the identification of 46 sequences – an acceptable number of sequences to enter into the next round of screening where homology models were created. However, potential trans-sialidases could easily have been lost, due to the specificity of the search motif and consequential risk of identifying false negatives. This risk was, however, accepted as a premise for obtaining a manageable quantity of sequences for homology modeling. False positives may have been included in the 46 identified sequences, due to the search motif aligning to a random site in one or more of the sequences, which is not unlikely given the size of the database and the short length of the search motif. Furthermore, since the initial sialidase database included putative enzymes, faulty classification of enzymes identified during the sequence screening would also result in identification of false positives. However, the presence of false positives amongst the identified sequences did not pose a problem since such sequences were highly unlikely to produce homology models with an intact active site. Such sequences were removed from the further analysis after failing the first modelling criterion.

Evaluation of homology modeling

Homology models for the 46 candidates were constructed and were, without regard to the quality of the models, aligned to the TcTS for inspection. Of the 46 sequences identified by the initial search, 13 produced enzyme candidate models displaying the correct active site alignment. In an effort to identify any sequences that may have been missed during the initial sequence search, the 13 sequences of the enzyme

candidate models were submitted to an NCBI BLAST search and close homologues were inspected to identify sequences with a correct Ω-R constellation. Such sequences were reintroduced to the study, modeled and evaluated as well. Two additional sequences were analyzed and number of candidates was ultimately adjusted to 15 sequences.

Selection of enzyme candidates for in vitro verification

Four candidates were selected for the *in vitro* verification study. The alignments of the four candidates with TcTS, leading to the selection are presented in Figure 2, while Table 3 summarizes the basis for the selection of each of the four candidates. The homology models of the selected candidates align correctly to the TcTS active site, displaying the correct amino acids, with two exceptions (SialM G296 and SialA T473). However, these two exceptions are likely the result of alignment errors by the modeling algorithm, since the expected amino acids were found in direct adjacency.

One candidate enzyme, SialH, fulfilled both predefined selection criteria; In addition to a perfectly aligned active site, the SialH model predicted the presence of an aromatic sandwich comprised by a tryptophan and a phenylalanine, aligning to the TcTS W312 and Y119, respectively (Table 3).

The remaining three candidates, SialP, SialA, and SialM, do not appear to harbor aromatic sandwiches. As predicted, the enzymes displayed an aromatic amino acid residues corresponding to TcTS W312, but the other half of the aromatic sandwich, corresponding to TcTS Y119 was missing (although in SialA a tryptophan was found adjacent to the residue G404 that aligns with TcTS_Tyr119, which could be a displacement modeling error). Although an aromatic sandwich was absent in the remaining enzymes, they were included in the study because of the closed, shielded nature of their binding clefts, which might facilitate the exclusion of water from the active site. The exclusion of water from the active site has previously been suggested to inhibit hydrolytic activity in engineered sialidases, albeit by a different suggested mechanism(Jers et al., 2014). Thus, a series of sialidase structure alignments were carried out to assert the notion that a closed active site will render trans-sialidase activity. The sialidases of *Vibrio colera*,

Clostridium perifringens and *Salmonella typhimurium* (VcSA, CpSA and StSA) have previously been reported to exhibit limited trans-sialidase activity(Schmidt, Sauerbrei, & Thiem, 2000) and they all displayed relatively closed active site structures compared to the sialidase of *Micromonospora viridifaciens* (MvSA), which is reported to exhibit no trans-sialidase activity(Smith & Eichinger, 1997). The alignment of the TcTS and MvSA is displayed in Figure 3 and the alignments of VcSA, CpSA and StSA to MvSA can be found in supplementary material.

Experimental characterization of trans-sialidase candidates

The four identified candidate enzymes were examined for trans-sialidase activity upon successful expression in *E. coli* as His-tagged proteins. As described previously, casein glycomacropeptide (cGMP) was used as donor substrate and lactose as the acceptor substrate for 3'SL production (Luo et al., 2014; Michalak et al., 2014; Nordvang et al., 2014) and this reaction was used in a time study setup, to evaluate trans-sialidase activity. CGMP is a side stream product from cheese production and contains 9 mM bound α-2,3- and α-2,6-linked sialic acid in a ratio of 1:1 (Villumsen et al., 2015), which makes it an interesting substrate for industrial production of sialyllated, high bulk compounds.The results of the trans-sialylation time studies are summarized in Figure 4 and Table 4. Our primary candidate, SialH, was capable of catalyzing trans-sialylation with a product to hydrolysis ratio peaking at above 2.5 for the production of 3'SL, making it the first bacterial trans-sialidase to be identified. The capability of this novel trans-sialidase rendered a relatively stable 3'SL concentration of 1.5 mM. Interestingly, another reaction product was formed besides 3'SL and SA. Analysed by MALDI-TOF, the product gave rise to an m/z 633-ion suggesting that it is a SL compound, but HPLC analysis ruled out that it was 6´SL (data not shown). The compound was susceptible to β-galactosidase hydrolysis, indicating that the compound is not sialylated at the non-reducing end (the galactose moiety). This was confirmed by NMR, and the novel compound was identified as 3-sialyllactose (3SL). 3'-sialylated glycans are the primary product of trans-sialylation reactions using the *Trypanosoma* trans-sialidases, but also 6'-sialylation has been reported(Schmidt et al., 2000). So far, 3SL is a

completely unknown trans-sialyllation product and it will be interesting to assess the functionality of the product in future.

For the three remaining enzyme candidates (selected based on the narrow acceptor site), only very limited trans-sialylation was observed and they primarily hydrolyzed the donor-bound SA, as seen from the rapid increase in free SA during the time course experiment. Product specificity varied among the candidates, which is not uncommon for this class of enzymes(Schmidt et al., 2000). SialP preferentially catalyzed synthesis of 3'SL but the novel product 3SL was also observed. SialM preferred synthesis of 3'SL over 6'SL whereas, in contrast, SialA appeared to exclusively synthesize 3'SL.

In case of SialH and SialA, it appears that concentrations of the different sialyllactose products reaches plateaus, but as would be expected for enzymes possessing hydrolase activity at the observed levels, all trans-glycosylation products were completely hydrolyzed after 48 h. The finding that SialH is the only trans-sialidase among the identified candidates supports the hypothesis that an aromatic sandwich above the active site is a suitable marker for identification of trans-sialidases.

NMR verification of reaction products

The ^{1}H NMR spectrum was used to elucidate the structure of the unknown sialylated lactose compound. The compound showed a signal at δ 4.537 (Gal H-1) ppm and a downfield shift of Glc H-3 (δ 3.56 → δ 3.73 ppm), suggested sialylation at O-3 of the Glc residue. Additionally, as there are no extra Neu5Ac axial or equatorial H-3 signals in the ^{1}H NMR spectrum, the Neu5Ac signals at δ 1.795 (Neu5Ac H-3a) and δ 2.734 (Neu5Ac H-3e) ppm also suggest the presence of a Neu5Ac (α2,3) residue. The combined evidence obtained in the NMR analysis confirmes that the novel trans-sialyllation product is 3SL.

In addition to structure elucidation of 3SL, the NMR analysis of the product sample was carried out to confirm the structures of the remaining product compounds and the ^{1}H NMR spectrum the reaction products confirmed the presence of 3'SL, 6'SL.

For 3'SL the ^{1}H NMR spectrum showed anomeric signals at δ 5.229 (Glc αH-1), δ 4.674 (Glc βH-1), and δ 4.535 (Gal H-1) ppm. The Neu5Ac signals at δ 1.795 (Neu5Ac H-3a) and δ 2.734 (Neu5Ac H-3e) ppm confirmed presence of a Neu5Ac (α2,3) residue. Furthermore, sialylation at O-3 of the Gal residue is apparent from the downfield shift of δ 3.663 to δ 4.111 (Gal H-3) ppm.

For 6'SL the ^{1}H NMR spectrum showed anomeric signals at δ 5.242 (Glc αH-1), δ 4.679 (Glc βH-1), and δ 4.48 (Gal H-1) ppm. The Neu5Ac signals at δ 1.707 (Neu5Ac H-3a) and δ 2.705 (Neu5Ac H-3e) ppm confirmed presence of a Neu5Ac (α2,6) residue.

Conclusions

The rational identification of novel trans-sialidases using a simple computational approach based on a putative acceptor binding sites was explored. From a database of 2909 sequences coding for genes categorized as sialidases, one candidate was predicted to have an active site displaying an aromatic sandwich comprised of a tryptophan and a phenylalanine (W366 and F168). Three additional candidates displayed a narrow binding cleft and were included in the further study. The candidates were expressed recombinantly and the purified enzymes were evaluated for their trans-sialidase activity. The enzyme with the (putative) aromatic sandwich binding site, SialH, was shown to exhibit a product to hydrolysis ratio of 2.5 (categorizing the enzyme as a trans-sialidase since $R_{P,H}>1$). The findings moreover underscored that Y119 and W312 are part of an aromatic sandwich structure that confers trans-sialylation activity in the TcTS. It is the first time that a sequence analysis approach has been successful in identifying a trans-sialidase and the first time a rational approach has led to the discovery of a trans-sialidase outside the *Trypanosoma* genus. The study was based on the hypothesis that an aromatic sandwich present immediately above the active site can be used as a trans-sialidase identification marker. The results of the study supported the initial hypothesis by SialH being the only trans-sialidase among the three analyzed candidates suspected to have very similar and relatively confining structures above their active sites.

Further analysis of the newly identified trans-sialidase of *H. parasuis* can provide a broader understanding of which characteristics provide an enzyme with trans-sialidase activity and thus facilitate the identification of more trans-sialidase enzymes. Surprisingly, a total of 3 sialylation products were identified during this study, including the novel trans-sialylation product 3SL which was produced efficiently by SialH ($R_{3SL,H}$=0.7). It is not guaranteed that the use of an aromatic sandwich as a trans-sialidase marker can be extrapolated to identification of other trans-glycosidases, but the possibility should be investigated. Furthermore, it is our suggestion that members of the glycobiology community investigate individual known trans-glycosidases in the search for useful identification markers such as the aromatic sandwich, which can lead to a breakthrough in trans-glycosidase identification.

References

Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., … Alzari, P. M. (2004). Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase. *Structure (London, England : 1993)*, *12*(5), 775–84. http://doi.org/10.1016/j.str.2004.02.036

Ammar, Z., Plazolles, N., Baltz, T., & Coustou, V. (2013). Identification of Trans-Sialidases as a Common Mediator of Endothelial Cell Activation by African Trypanosomes. *PLoS Pathogens*, *9*(10), e1003710. http://doi.org/10.1371/journal.ppat.1003710

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. http://doi.org/10.1093/nar/28.1.235

Bissaro, B., Monsan, P., Fauré, R., & O'Donohue, M. J. (2015). Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *The Biochemical Journal*, *467*(1), 17–35. http://doi.org/10.1042/BJ20141412

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*, *22*(9), 1147–62. http://doi.org/10.1093/glycob/cws074

Buscaglia, C. A., Campo, V. A., Frasch, A. C. C., & Di Noia, J. M. (2006). Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nature Reviews. Microbiology*, *4*(3), 229–36. http://doi.org/10.1038/nrmicro1351

Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., & Alzari, P. M. (2002). The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis. *Molecular Cell*, *10*(4), 757–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12419220

Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., París, G., … Alzari, P. M. (2000). Structural basis of sialyltransferase activity in trypanosomal sialidases. *The EMBO Journal*, *19*(1), 16–24. http://doi.org/10.1093/emboj/19.1.16

Desmet, T., Soetaert, W., Bojarová, P., Křen, V., Dijkhuizen, L., Eastwick-Field, V., & Schiller, A. (2012). Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, *18*(35), 10786–801. http://doi.org/10.1002/chem.201103069

Holck, J., Larsen, D. M., Michalak, M., Li, H., Kjærulff, L., Kirpekar, F., … Meyer, A. S. (2014). Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by Trypanosoma cruzi trans-sialidase. *New Biotechnology*, *31*(2), 156–65. http://doi.org/10.1016/j.nbt.2013.11.006

Jers, C., Michalak, M., Larsen, D. M., Kepp, K. P., Li, H., Guo, Y., … Mikkelsen, J. D. (2014). Rational design of a new Trypanosoma rangeli trans-sialidase for efficient sialylation of glycans. *PloS One*, *9*(1), e83902. http://doi.org/10.1371/journal.pone.0083902

KOSHLAND, D. E. (1953). STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews*, *28*(4), 416–436. http://doi.org/10.1111/j.1469-185X.1953.tb01386.x

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(Database issue), D490–5. http://doi.org/10.1093/nar/gkt1178

Luo, J., Nordvang, R. T., Morthensen, S. T., Zeuner, B., Meyer, A. S., Mikkelsen, J. D., & Pinelo, M. (2014). An integrated membrane system for the biocatalytic production of 3'-sialyllactose from dairy by-products. *Bioresource Technology*, *166C*, 9–16. http://doi.org/10.1016/j.biortech.2014.05.003

Michalak, M., Larsen, D. M., Jers, C., Almeida, J. R. M., Willer, M., Li, H., … Mikkelsen, J. D. (2014). Biocatalytic production of 3'-sialyllactose by use of a modified sialidase with superior trans-sialidase activity. *Process Biochemistry*, *49*(2), 265–270. http://doi.org/10.1016/j.procbio.2013.10.023

Montagna, G., Cremona, M. L., Paris, G., Amaya, M. F., Buschiazzo, A., Alzari, P. M., & Frasch, A. C. C. (2002). The trans-sialidase from the african trypanosome Trypanosoma brucei. *European Journal of Biochemistry / FEBS*, *269*(12), 2941–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12071958

Nordvang, R. T., Luo, J., Zeuner, B., Prior, R., Andersen, M. F., Mikkelsen, J. D., … Pinelo, M. (2014). Separation of 3'-sialyllactose and lactose by nanofiltration: A trade-off between charge repulsion and pore swelling induced by high pH. *Separation and Purification Technology*, *138*, 77–83. http://doi.org/10.1016/j.seppur.2014.10.012

Osorio, L., Ríos, I., Gutiérrez, B., & González, J. (2012). Virulence factors of Trypanosoma cruzi: who is who? *Microbes and Infection / Institut Pasteur*, *14*(15), 1390–402. http://doi.org/10.1016/j.micinf.2012.09.003

Paris, G., Ratier, L., Amaya, M. F., Nguyen, T., Alzari, P. M., & Frasch, A. C. C. (2005). A sialidase mutant displaying trans-sialidase activity. *Journal of Molecular Biology*, *345*(4), 923–34. http://doi.org/10.1016/j.jmb.2004.09.031

Rather, M., & Mishra, S. (2013). β-Glycosidases: An alternative enzyme based method for synthesis of alkyl-glycosides. *Sustainable Chemical Processes*, *1*(1), 7. http://doi.org/10.1186/2043-7129-1-7

Schenkman, R. P., Vandekerckhove, F., & Schenkman, S. (1993). Mammalian cell sialic acid enhances invasion by Trypanosoma cruzi. *Infection and Immunity*, *61*(3), 898–902. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=302817&tool=pmcentrez&rendertype=abstract

Schmidt, D., Sauerbrei, B., & Thiem, J. (2000). Chemoenzymatic synthesis of sialyl oligosaccharides with sialidases employing transglycosylation methodology. *The Journal of Organic Chemistry*, *65*(25), 8518–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11112571

Smith, L. E., & Eichinger, D. (1997). Directed mutagenesis of the Trypanosoma cruzi trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. *Glycobiology*, *7*(3), 445–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9147054

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. http://doi.org/10.1016/j.pep.2005.01.016

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–8. http://doi.org/10.1093/nar/gki408

Tiralongo, E., Martensen, I., Grötzinger, J., Tiralongo, J., & Schauer, R. (2003). Trans-sialidase-like sequences from Trypanosoma congolense conserve most of the critical active site residues found in other trans-sialidases. *Biological Chemistry*, *384*(8), 1203–13. http://doi.org/10.1515/BC.2003.133

Villumsen, N. S., Jensen, H. B., Thu Le, T. T., Møller, H. S., Nordvang, R. T., Nielsen, L. R., … Larsen, L. B. (2015). Self-assembly of caseinomacropeptide as a potential key mechanism in the formation of visible storage induced aggregates in acidic whey protein isolate dispersions. *International Dairy Journal*, *49*, 8–15. http://doi.org/10.1016/j.idairyj.2015.05.003

Watson, J. N., Newstead, S., Dookhun, V., Taylor, G., & Bennet, A. J. (2004). Contribution of the active site aspartic acid to catalysis in the bacterial neuraminidase from Micromonospora viridifaciens. *FEBS Letters*, *577*(1-2), 265–9. http://doi.org/10.1016/j.febslet.2004.10.016

Zeuner, B., Jers, C., Mikkelsen, J. D., & Meyer, A. S. (2014). Methods for improving enzymatic trans-glycosylation for synthesis of human milk oligosaccharide biomimetics. *Journal of Agricultural and Food Chemistry*, *62*(40), 9615–31. http://doi.org/10.1021/jf502619p

Zeuner, B., Luo, J., Nyffenegger, C., Aumala, V., Mikkelsen, J. D., & Meyer, A. S. (2014). Optimizing the biocatalytic productivity of an engineered sialidase from Trypanosoma rangeli for 3'-sialyllactose production. *Enzyme and Microbial Technology*, *55*, 85–93. http://doi.org/10.1016/j.enzmictec.2013.12.009

Aasland, R., Abrams, C., Ampe, C., Ball, L. J., Bedford, M. T., Cesareni, G., … Winder, S. J. (2002). Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Letters*, *513*(1), 141–144. http://doi.org/10.1016/S0014-5793(01)03295-1

1 **Tables:**

2 **Table 1:** Summary of the interactions between the TcTS and 3'SL in the MC.

TcTS properties	Residues	Role of residues	Hydrogen bonds (to substrate)	Hydrophobic interactions	Strictly conserved
Sialic acid moiety fixation	Arg35	Carboxylate fixation	SA_O1A, SA_ O1B	-	+
	Arg245			-	+
	Arg314			-	+
	Asp96	Acetamide fixation	SA_N5	-	+
	Trp120	Glycerol fixation	SA_O9	-	+
	Gln195		SA_O8, SA_O9	-	
	Arg53	Ring fixation	SA_O4	-	+
Catalysis	Asp59	Acid/base catalyst	SA_O2	-	+
	Tyr342	Enzymatic nucleophile	(covalent bond to C2)	-	+
Lactose moiety fixation	Tyr119	Aromatic sandwich	-*	+	-
	Trp312		-	+	-

3 *The Tyr119 goes through a conformational change during the catalytic process, where a hydrogen bond to the substrate is
4 formed(Amaya et al., 2004).

5 **Table 2:** An overview of the origin and selected properties of the candidate genes

Candidate	**Organism name**	**Uniprot ID**	**Submitted Name**	**Mw**	**Length**
SialA	*Actinomyces oris Actinomyces viscosus**	Q44562*	NanH Sialidase*	73,089	685 AA
SialH	*Haemophilus parasuis*	I7AJQ9	NanH Sialidase	89,948	803 AA
SialM	*Manheimia heamolytica*	E2NYK4	NanH Sialidase	89,245	795 AA
SialP	*Pasteurella multocida*	I1VL53	NanH Sialidase	89,217	798 AA

6 *closest BLAST alignment – original sequence has been removed from database.

7 **Table 3:** Summary of the 3D alignment visualized in Figure 2.

TcTS AC properties	TcTS	SialH	SialP	SialA	SialM
Sialic acid moiety fixation	R35	R80	R73	R311	R79
	R245	R298	R291	R524	~~*G296*~~ (R297)[1]
	R314	R368	R359	R592	R360
	D96	D143	D136	D379	D142
	W120	W169	W162	W403	W168
	Q195	Q245	Q238	~~*T473*~~ (Q472)[1]	Q243
	R53	R99	R92	R330	R98
Catalysis	D59	D105	D98	D340	D104
	Y342	Y402	Y393	Y620	Y394
Lactose moiety fixation	Y119	***F168***[3]	*N161*[2]	~~*G404*~~ (W403)[1]	*Q167*[2]
	W312	W366	***Y357***[3]	W590	W358

8 [1] Modelling error suspected - alternative alignment in brackets, [2] Non-aromatic residue, [3] Alternative aromatic residue

9

Table 4: Best $R_{P,H}$ for the four candidates(over the time course).

		3'SL	6'SL	3SL
$R_{P,H}$	**SialH**	2,5	0,010	0,70
	SialP	0,066	0,004	0,020
	SialA	n.a.	0,016	n.a.
	SialM	0,088	0,020	n.a.

Figures:

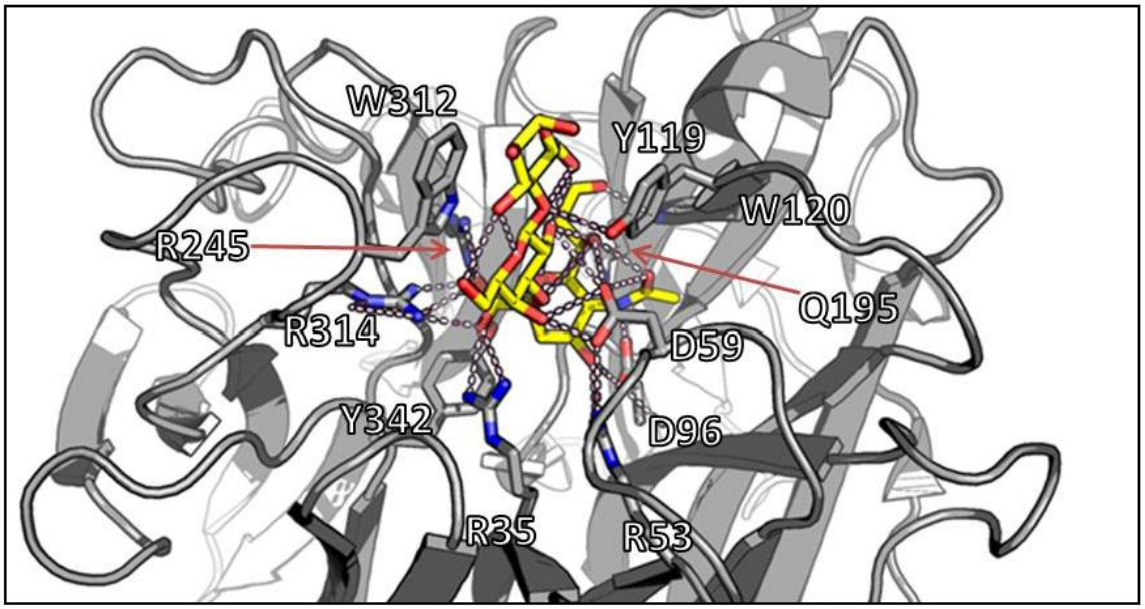


Fig. 1

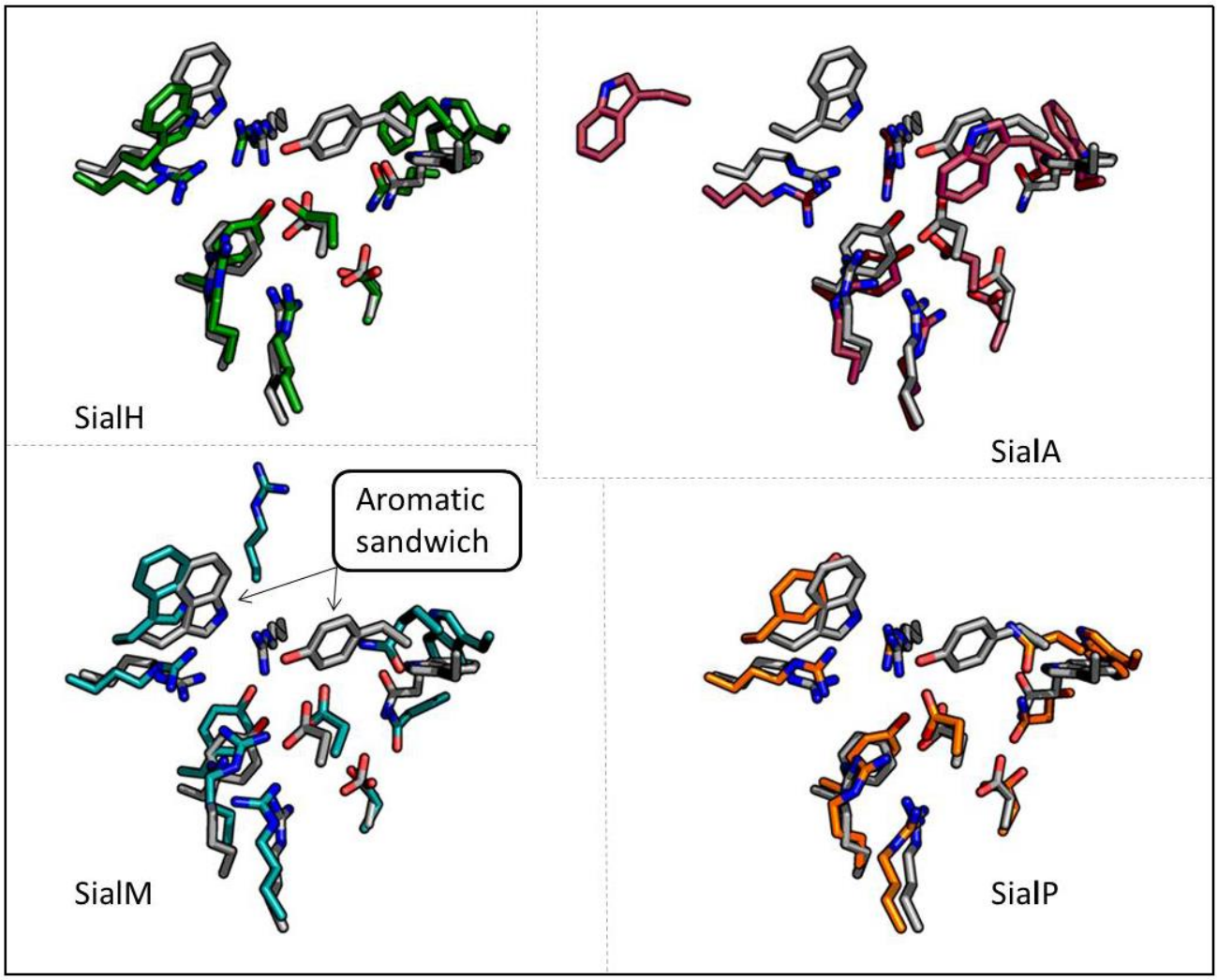


Fig. 2

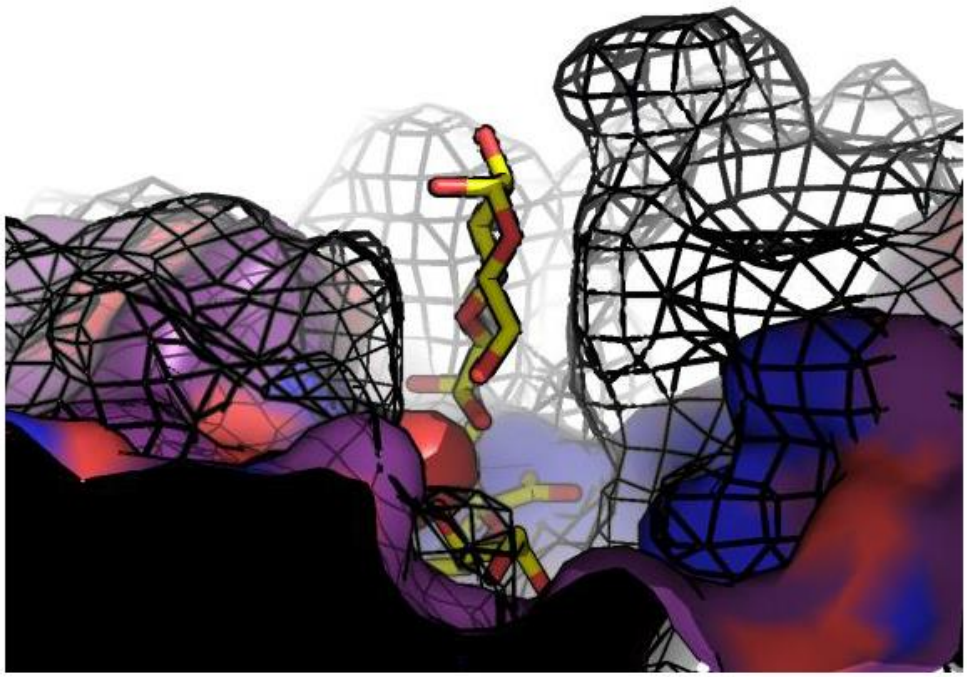


Fig. 3

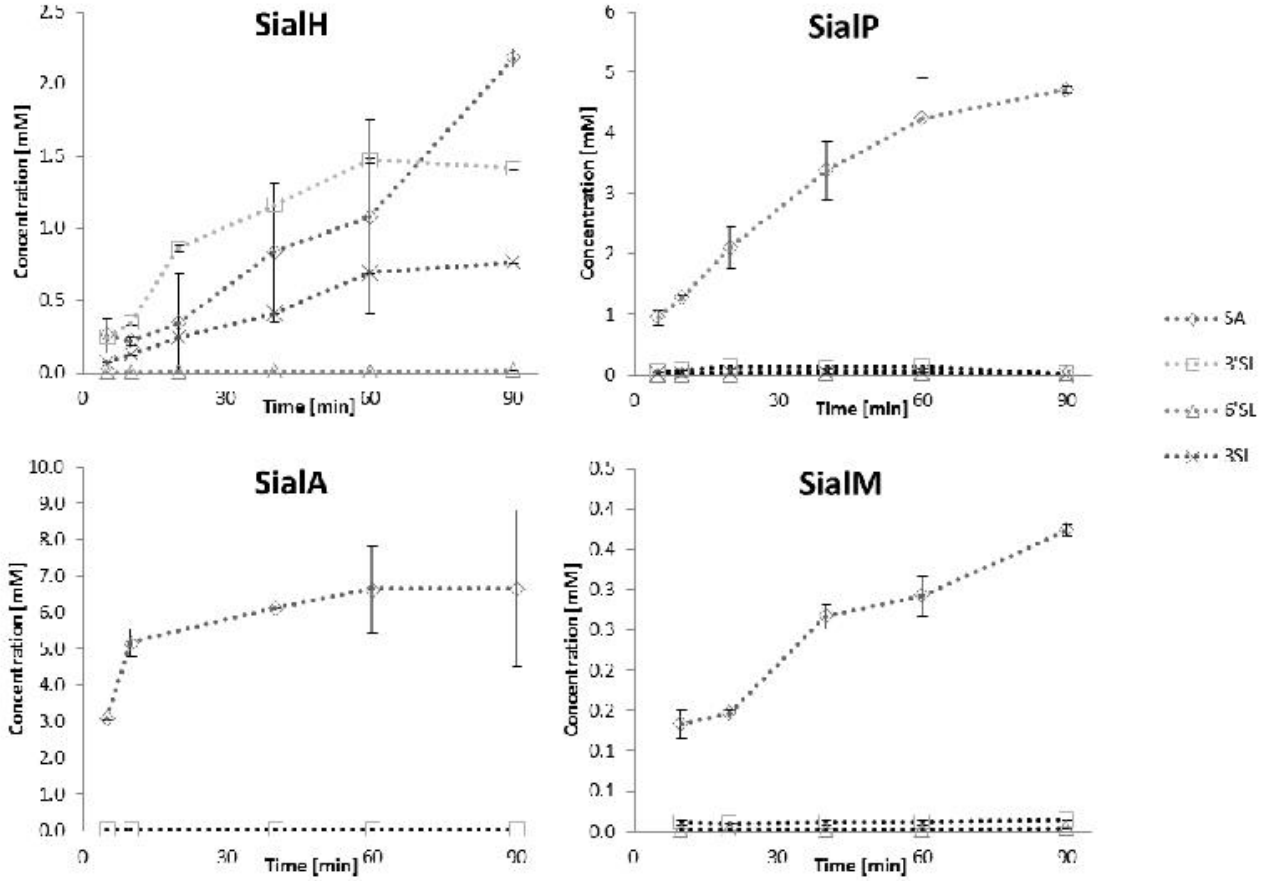
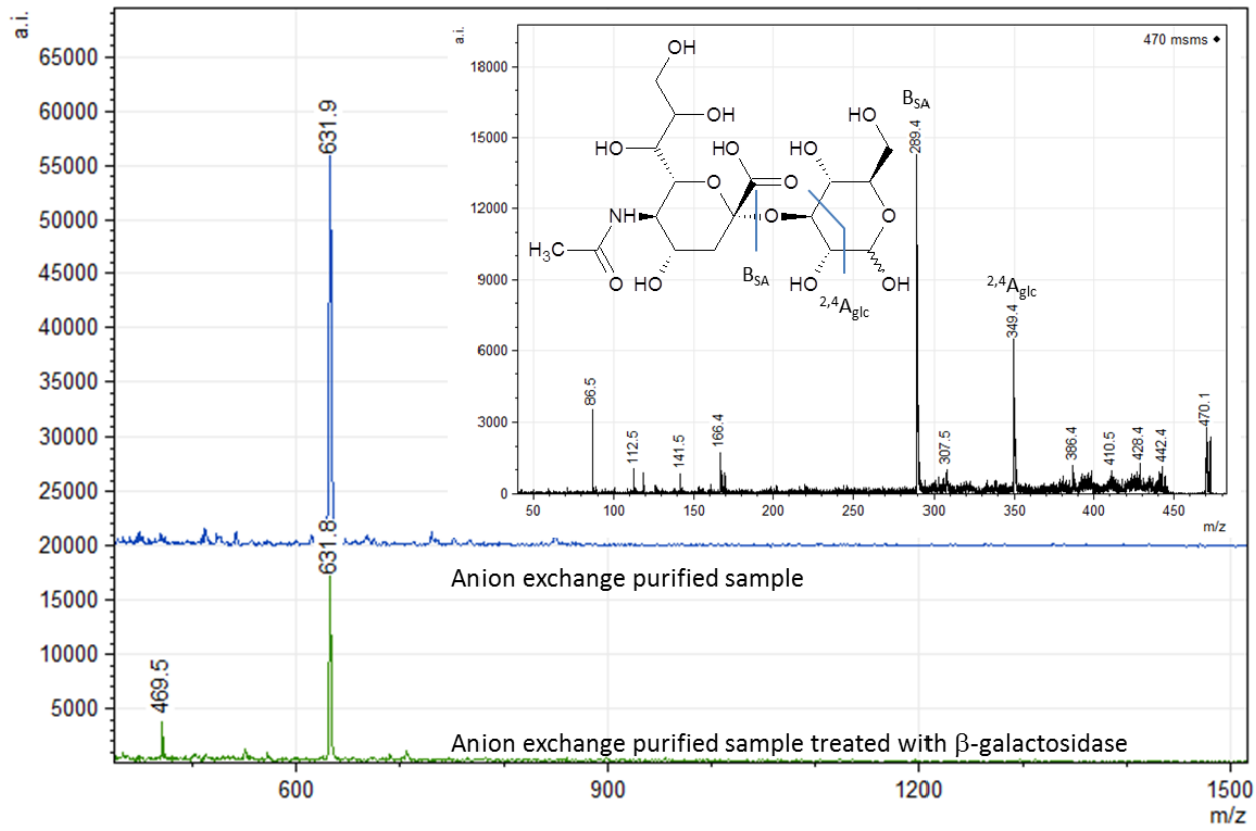

SialH
SialP
SialA
SialM
Concentration [mM]
Time [min]
SA
3'SI
6'SL
3SI

Fig. 4

1


a.i.
631.9
631.8
469.5
Anion exchange purified sample
Anion exchange purified sample treated with β-galactosidase
600
900
1200
1500
m/z
470 msms
289.4
349.4
86.5
112.5
141.5
166.4
307.5
366.4
410.5
428.4
442.4
470.1
BSA
2,4Aglc

2

3 **Fig. 5**

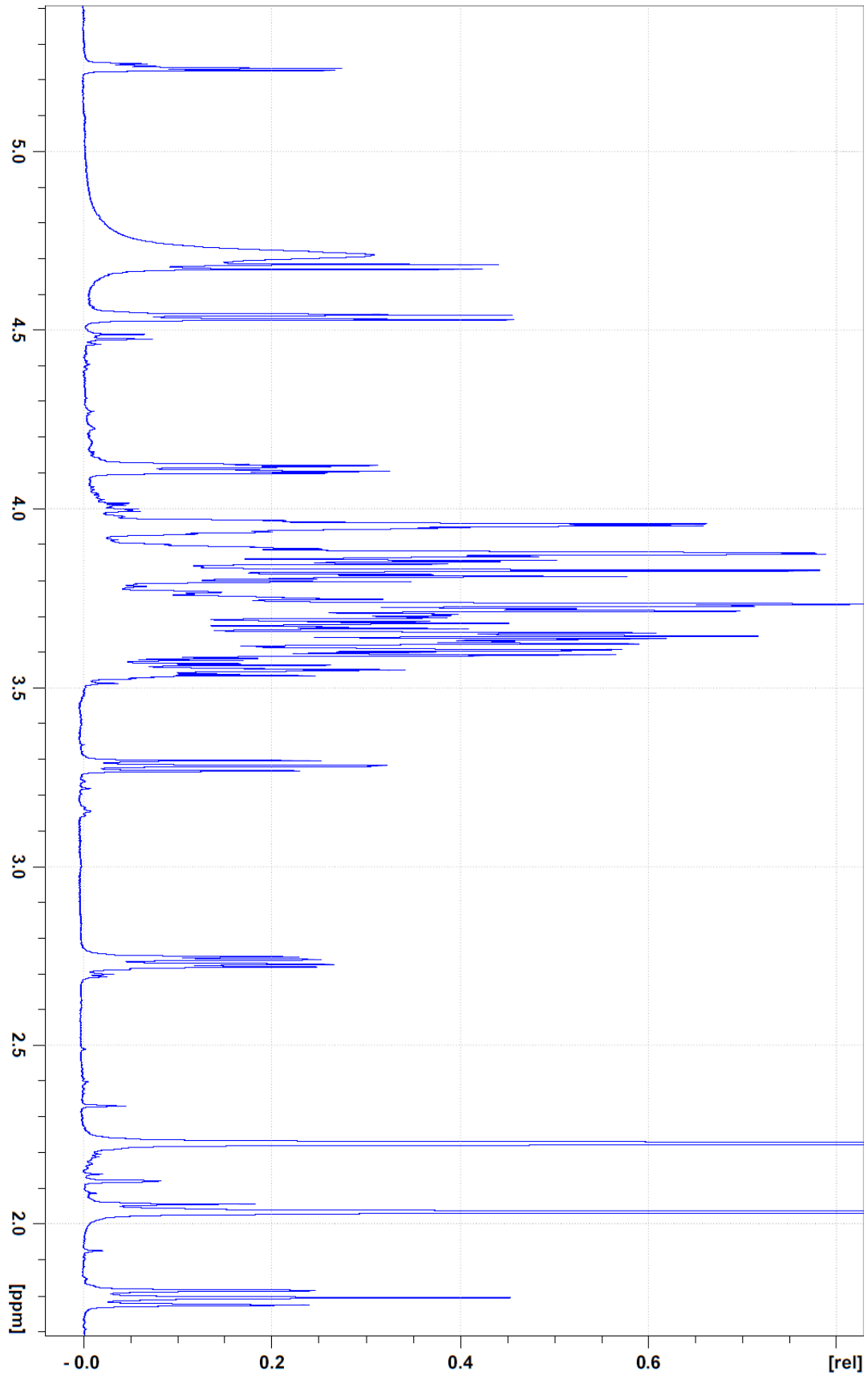


Fig. 6

Figure Legends:

Fig. 1 The 3'SL substrate (yellow) is shown positioned in the active site of the TcTS. The active site residues (the top 9 residues in Table 1) bind the SA moiety of the substrate by a number of hydrogen bonds (A), whereas the aromatic sandwich (Trp312 and Tyr119) align to the lactose moiety and bind it in the tight cleft above the active site (B). An overview of the interactions between enzyme and substrate is given in Table 1.

Fig. 2 The four enzyme candidates selected for *in vitro* testing; SialH (green), SialP (orange), SialA (red) and SialM (blue) were modeled by homology modeling using the TcTS as the template (gray). The active site and aromatic sandwich residues are also shown in this figure (the aromatic sandwich is indicated with arrows in the SialM figure). The four candidates were selected based on two criteria 1) restoration of the active site with its 9 conserved residues and 2) the formation of a narrow (closed) binding cleft above the active site, preferably with an aromatic sandwich composed of two aromatic amino acid residues. Whereas all of the selected candidates were predicted to have sterically closed active sites, SialH was the only candidate predicted to display an aromatic sandwich. The aromatic sandwich of SialH is comprised of a tryptophan and a phenylalanine and aligns (in the model) almost perfectly to the one found in the TcTS.

Fig. 3 The figure shows the alignment of the surfaces around the active sites of the TcTS (balck mesh surface) and the *M. viridifaciens* sialidase (magenta surface) in MC with 3'SL (yellow sticks). The 3'SL is relatively shielded in the TcTS compared to the MvSA and it seems convincing that water will more easily be able to attack the Cl of MvSA. Therefore, it is suggested that candidate enzymes should have a closed active site because 1) the identification of a trans-sialidase seem more likely if the active site is shielded and 2) to support the hypothesis that an aromatic sandwich is essential to trans-sialidase activity if assessed candidates with a shielded active site show no trans-sialidase activity.

Fig. 4 The four candidate enzymes were used to catalyze reactions using CGMP as SA donor and lactose as acceptor. Formation of reaction products SA (all enzymes), 3'SL (SialH and SialP), 6'SL (SialP, SialA and SialM) and the novel trans-sialidase reaction product 3SL (SialH and SialP) were monitored over a period of 90 minutes and it is clear from the graphs that only SialH, which was the candidate with an aromatic sandwich, displayed a positive product to hydrolysis ration. Thus SialH can be classified as a trans-sialidase whereas the remaining candidates are primarily hydrolytic sialidases.

Fig. 5 MALDI-TOF of anion exchange purified HN reaction mixture, before and after β-galactosidase treatment. MALDI-TOF/TOF of 632 in the untreated sample is consistent with the presence of a SL type compound. MALDI-TOF/TOF of m/z 470 ion in the β-galactosidase treated sample is the result of sialyllactose breakdown, which suggests a non-sialylalted galactose moiety and thus, a sialyllated glucose moiety. MS-MS of the m/z 470 ion is consistent with breakdown of sialylalted glucose. Two ions resulted in significant peaks (BSA and Aglc) and are, together with the suggested bond cleavage, indicated in the figure.

Fig. 6

To confirm the obtained reaciton products, samples wre submitted to NMR analysis. As expected, the ^{1}H NMR spectrum of the reaction products reveals the presence of 3'-sialyllactose (3'SL) and 6'-sialyllactose (6'SL). Of special intrest was the signal at δ 4.537 (Gal H-1) ppm, rendered by the unknown reaction product. Additionally a possible downfield shift of Glc H-3 (δ 3.56 → δ 3.73 ppm) suggested sialylation at O-3 of the Glc residue. Since there are no extra Neu5Ac axial or equatorial H-3 signasl in the ^{1}H NMR spectrum, the Neu5Ac signals at δ 1.795 (Neu5Ac H-3a) and δ 2.734 (Neu5Ac H-3e) ppm also suggest an (α2,3)-linked Neu5Ac residue. Thus, the unknown sialyllation product was identified as 3-sialyllactose (3SL).

Chapter 3: Backbone structures in human milk oligosaccharides: trans-glycosylation by metagenomic β-N-acetylhexosaminidases

3.1 Motivation

While the main focus of this PhD project has been 3'SL production, another goal of the overall HMO project (of which this PhD study was a part), was to produce a wider range of HMOs. Thus it was necessary to develop the means to produce the backbone structures of the HMOs which as described are comprised of lacto-N-biose and N-acetyllactosamine. The strategy for HMO backbone production was predefined with the first step being elongation of lactose by a GlcNAc unit. The motivation for this strategy was partly the availability of cheap donor source as well as the suspicion that members of the otherwise hydrolytic β-N-acetylhexosaminidase enzyme family (GH 20) would be able to efficiently carry out transfer of the GlcNAc to an acceptor molecule (as it has been shown to be the case for other GH type enzymes).

3.2 Hypotheses and objectives

An objective of this study was to identify novel β-N-acetylhexosaminidases and to evaluate trans-glycosylation activity of any such enzymes. Thus the following hypotheses were stated.

Hypothesis 3.1: Because GlcNAc in the polymeric form of chitin as exoskeleton in many lower eukaryotes such as insects and arachnids inhabiting soil, occurrence of GHs capable of degrading GlcNAc structures are expected to be possible to identify from a soil derived metagenomics library.

Hypothesis 3.2: Among GH20 enzymes identified from metagenomics libraries, specimens can be found which are able to transfer a GlcNAc residue to various acceptor molecules.

3.3 Experimental considerations

Since no enzymes with exclusive trans-hexosaminidase activity has been discovered it was decided that the best strategy for enzyme identification in this case was a screening for hydrolytic enzymes with subsequent assessment of trans-activity by a method applicable to the quantity of identified hydrolases.

For the initial hydrolase identification it was decided to screen a metagenomic library. As described the use of metagenomics libraries have many advantages. However the choice of metagenomics library, used for this screening, might seem strange as its origin was from soil. Although soil derived metagenomics libraries have successfully been used for identification of GH family enzymes, the main reason for choosing this library was that it had already been constructed (for a different purpose) and was readily available. It can thus be speculated that better results could have been obtained if a genomic library had been constructed from a source where hexosaminidase activity would be expected to overrepresented. However it is not immediately obvious where such an environment can be found and the metagenomics library at hand was used, improving the return of investment on the resources already put into its construction.

In this study the screening of the metagenomic library was hampered by the limitations on reporter-substrate. The substrate used was X-GlcNAc and thus it was only possible to identify clones expressing enzymes capable of degrading this substrate. Thus it was also here necessary to use a natural substrate for enzymatic reactions followed by HPLC analysis. Fortunately the trans-glycosylation results were good, but no such outcome was predictable. The amounts of experimental work that went into this screening, once again emphasize the need for a high throughput trans-glycosylation screening method.

The characterization of the enzymes were done using the synthetic substrates pNP-GlcNAc and pNP-GalNAc (pNP-N-acetylgalactosamine), which proved troublesome as the substrates were very unstable in standard laboratory conditions. Much work was undertaken to develop a suitable assay, which in the end relied primarily on speedy execution to achieve good results. However, it is worth noting that 1) these substrates are different from the natural substrates and that 2) these results only reflect the hydrolytic activity of the enzymes.

No attempts were made to optimize the reaction conditions for the trans-glycosylation reaction as the executed study with its scope produces a comprehensive publication. Furthermore suggested strategies for reaction optimization may include product removal which would have taken the publication in a different direction.

3.4 Conclusions

By the means of metagenomics library screening on X-GlcNAc loaded agar plates two novel β-N-acetylhexosaminidases HEX1 and HEX2 were identified. In enzymatic reaction setups with a natural donor- and acceptor-substrates (chitin oligomers and lactose respectively) the enzymes proved capable of synthesizing an HMO backbone precursor structure (lacto-N-triose) with fair yields (2 % and 8 % on donor for Hex1 and Hex2 respectively). Thus product concentrations reached levels of up to 8 mM which might be applicable for industrial use depending on efficiency of product purification. Additionally GlcNAc transfer to sucrose, maltose, galactose, and glucose was also successfully carried out by both enzymes.

3.5 Paper II

Appl Microbiol Biotechnol (2015) 99:7997–8009
DOI 10.1007/s00253-015-6550-0

Backbone structures in human milk oligosaccharides: trans-glycosylation by metagenomic β-*N*-acetylhexosaminidases

Christian Nyffenegger[1] · Rune Thorbjørn Nordvang[1] · Birgitte Zeuner[1] · Mateusz Łężyk[1] · Elisabetta Difilippo[2] · Madelon J. Logtenberg[2] · Henk A. Schols[2] · Anne S. Meyer[1] · Jørn Dalgaard Mikkelsen[1]

Received: 6 January 2015 / Revised: 13 March 2015 / Accepted: 16 March 2015 / Published online: 7 April 2015

Abstract This paper describes the discovery and characterization of two novel β-*N*-acetylhexosaminidases HEX1 and HEX2, capable of catalyzing the synthesis of human milk oligosaccharides (HMO) backbone structures with fair yields using chitin oligomers as β-*N*-acetylglucosamine (GlcNAc) donor. The enzyme-encoding genes were identified by functional screening of a soil-derived metagenomic library. The β-*N*-acetylhexosaminidases were expressed in *Escherichia coli* with an N-terminal His_6-tag and were purified by nickel affinity chromatography. The sequence similarities of the enzymes with their respective closest homologues are 59 % for HEX1 and 51 % for HEX2 on the protein level. Both β-*N*-acetylhexosaminidases are classified into glycosyl hydrolase family 20 (GH 20) are able to hydrolyze *para*-nitrophenyl-β-*N*-acetylglucosamine (*p*NP-GlcNAc) as well as *para*-nitrophenyl-β-*N*-acetylgalactosamine (*p*NP-GalNAc) and exhibit pH optima of 8 and 6 for HEX1 and HEX2, respectively. The enzymes are able to hydrolyze *N*-acetylchitooligosaccharides with a degree of polymerization of two, three, and four. The major findings were, that HEX1 and HEX2 catalyze trans-glycosylation reactions with lactose as acceptor, giving rise to the human milk oligosaccharide precursor lacto-*N*-triose II (LNT2) with yields of 2 and 8 % based on the donor substrate. In total, trans-glycosylation reactions were tested with the disaccharide acceptors β-lactose, sucrose, and maltose, as well as with the monosaccharides galactose and glucose resulting in the successful attachment of GlcNAc to the acceptor in all cases.

Keywords Functional screening · Protein expression · Synthetic biology · Chito-oligosaccharides · Lacto-*N*-triose II

Introduction

Human milk oligosaccharides (HMO) are abundant in human breast milk and reach concentrations up to 20 g/L (Coppa et al. 1999; Gabrielli et al. 2011). Breast-fed neonates benefit from the HMO having (i) a prebiotic function, notably promoting growth of *Bifidobacterium longum* subsp. *infantis*, (ii) antiadhesive action towards pathogenic microorganisms, (iii) regulation of immune cell response, and (iv) ability to lower the risk for necrotizing enterocolitis (Bode 2009, 2012). HMOs are composed of five different monosaccharides and share a basic core structure with β-lactose at the reducing end. To synthesize HMOs with a degree of polymerization higher than three, β-lactose is primarily elongated by β-*N*-acetylglucosamine (GlcNAc), linked via a β-1,3-glycosidic bond to yield the trisaccharide lacto-*N*-triose II (LNT2; GlcNAcβ1-3Galβ1-4Glc), a HMO backbone precursor. Alternatively, β-lactose can be elongated by GlcNAc forming a β-1,6-glycosidic bond (Kunz et al. 2000; Boehm and Stahl 2007). In mammary glands, the synthesis of LNT2 may start with the transfer of an activated GlcNAc residue to the galactose (Gal) residue of β-lactose by a β-*N*-acetylglucosaminyltransferase (Kobata 2003). LNT2 can then be elongated with Gal to yield the tetrameric HMO

Electronic supplementary material The online version of this article (doi:10.1007/s00253-015-6550-0) contains supplementary material, which is available to authorized users.

✉ Jørn Dalgaard Mikkelsen
jdm@kt.dtu.dk

[1] Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, Søltofts Plads, 2800 Kgs. Lyngby, Denmark

[2] Laboratory of Food Chemistry, Wageningen University, Bornse Weilanden 9, 6708 Wageningen, The Netherlands

structures lacto-*N*-tetraose (LNT) or lacto-*N*-neotetraose (LNnT) (Bode 2012). Further elongation of the backbone is achieved through the alternating addition of GlcNAc and Gal at the non-reducing end yielding the structure [Gal(β1-3/4)GlcNAc(β1-]$_n$ 3/6) with $n \leq 25$ (Kunz et al. 2000; Boehm and Stahl 2007). The prebiotic function of HMO based on LNT2 is presumed to be associated with the growth of, e.g., *Bifidobacteria* (Miwa et al. 2010; Nguyen et al. 2012), serving as precursors in bacterial cell wall synthesis (McVeagh and Miller 1997).

Biocatalytic or chemical synthesis of HMO and HMO-like molecules with a degree of polymerization (DP) of four or higher require efficient elongation by attaching GlcNAc to galactose or β-lactose. Several strategies have been applied to produce the trisaccharide LNT2 or to elongate a terminal Gal in an oligosaccharide by a GlcNAc residue. Chemical synthesis requires a multistep approach including protection, glycosylation, and deprotection in order to avoid unwanted side reactions (Kościelak et al. 1979; Ito et al. 1988; Kameyama et al. 1990). Enzyme-based approaches were tested using β-*N*-acetylglucosaminyltransferases (Murata et al. 1999; Blixt et al. 1999) or β-*N*-acetylhexosaminidases (Matahira et al. 1995; Singh et al. 1995; Murata et al. 1997; Nilsson et al. 1999). These reactions rely on nucleotide-activated or *p*NP-modified GlcNAc donors or non-natural saccharide derivatives. However, activated donor compounds or non-natural acceptors are expensive in their synthesis, may not be food-grade classified, and are thus not suited for a cost-efficient HMO production. Matsuo and coworkers used the inexpensive substrates GlcNAc and β-lactose in combination with a recombinant β-*N*-acetylglucosaminidase from *Aspergillus oryzae* to produce LNT2 by trans-glycosylation with a reaction yield of 0.21 % (Matsuo et al. 2003). In order to achieve higher biocatalytical yields of LNT2 via trans-glycosylation, other cheap GlcNAc-containing substrates may be better suited, e.g., chitin derivatives from side streams of the marine fish industry (shells from crustaceans).

A first prerequisite for an enzyme to use chitin derivatives as GlcNAc donor compound is its ability to catalyze the cleavage of the donor. In order to find promising enzyme candidates capable of utilizing chitin derivatives, activity screenings of metagenomic libraries can be applied, e.g., by visualizing the hydrolysis of the putative donor-substrate in the form of 5-bromo-4-chloro-3-indolyl-GlcNAc. Metagenomic libraries of microbial soil communities generally contain a multitude of genes coding for various enzyme classes and were successfully used in identifying novel oxidoreductases (Knietsch et al. 2003), lipases (Henne et al. 2000), and glycosyl hydrolases (Beloqui et al. 2010). The advantages of using these libraries are (i) the access to enzyme-coding genes from rare or uncultivable bacteria, which constitute the major part of soil microorganisms (Torsvik et al. 1990) and (ii) the discovery of novel enzymes with low similarity to known enzymes (Ferrer et al. 2005; Wang et al. 2009; Findley et al. 2011). The frequency of soil-derived metagenomic library clones expressing an enzyme with a desired activity is generally low. The chance of finding an enzyme capable of forming trans-glycosylation products from GlcNAc-containing compounds (originating from chitin shells of crustaceans) in soil metagenomic libraries can be increased by choosing soil where GlcNAc-containing components are abundant, e.g., dump sites. In combination with the library construction, simple but sensitive high-throughput activity-based screening methods are required to (i) be able to screen a large number of clones and (ii) identify genes coding for β-*N*-acetylhexosaminidases with low homology to known genes.

In the present study, we identified and characterized two novel β-*N*-acetylhexosaminidases which have the capacity to synthesize HMO backbone precursors such as LNT2 using lactose as starting molecule.

Materials and methods

Chemicals

5-Bromo-4-chloro-3-indolyl *N*-acetyl-b-D-glucosaminide (X-GlcNAc), 4-nitrophenyl-*N*-acetyl-β-D-glucosamine (*p*NP-GlcNAc), 4-nitrophenyl *N*-acetyl-β-D-galactosamine (*p*NP-GalNAc), *N*,*N*′-diacetylchitobiose, *N*,*N*′,*N*″-triacetylchitotriose, *N*,*N*′,*N*″,*N*‴-tetraacetylchitotetraose, and lacto-*N*-triose II (LNT2) standard were purchased from Carbosynth (Compton, UK). Restriction enzymes were purchased from Thermo Fischer Scientific (Massachusetts, USA). Chitin (from shrimp shells), chitosan (with a molecular weight between 60 and 120 kDa), β-lactose, sucrose, D(+)-maltose, D(+)-galactose, D(+)-glucose, β-*N*-acetylglucosamine, ampicillin, chloramphenicol, and imidazole were purchased from Sigma-Aldrich (Steinheim, Germany).

Strains and plasmids

For the maintenance and propagation of a constructed soil metagenomic library, a Δ*lacZYA* derivative of the EPI300™-T1^R strain (Epicentre Biotechnologies, USA) was used, which was transformed with the heat-inducible lysis vector pEAS-1a (DualSystems Biotech-Switzerland, Schlieren). The *Escherichia coli* strain DH5α was used for subcloning and plasmid propagation and BL21 (DE3), C41 (DE3), and C43 (DE3) (Novagen and Lucigen, USA) were used for recombinant production of β-*N*-acetyl-hexosaminidases. The plasmids used for subcloning and expression were pUC18 (Thermo Scientific, USA) and pETM-10 (EMBL), respectively.

Construction of metagenomic library

A 10-g soil sample was obtained from Mining and Recycling Company RGS 90 A/S (Copenhagen). Metagenomic DNA was extracted from the soil with the PowerMax Soil DNA Isolation kit (Mobio Laboratories Inc.), following the manufacturer's recommendations. *E. coli*-based libraries of 40–50 kb metagenomic DNA were created using the CopyControl Fosmid Library Production kit (Epicentre Biotechnologies, USA). DNA was end-repaired, and high molecular weight (40-50 kb) fragments were size selected and purified using overnight 1 % low melting-point agarose gel electrophoresis; 0.25 μg of metagenomic DNA was ligated with 0.5 μg of the linearized fosmid pCC1FOS vector and packaged using replication-deficient phage extract. The Δ*lacZYA* derivative of the EPI300™-T1^{R} strain, transformed with the heat-inducible lysis vector pEAS-1a was transfected with the resulting phages, and the size of the library was determined by plating serial dilutions on lysogeny broth (LB) agar plates containing 12.5 μg/mL chloramphenicol and 100 μg/mL ampicillin. *E. coli* metagenomic DNA library was grown to mid-log phase for 8 h shaking at 30 °C in 50 mL LB supplemented with relevant antibiotics. Two-milliliter aliquots with 50,000 colony forming units/mL were stored frozen at −80 °C in the presence of 15 % glycerol. An average redundancy of about one to two clones per library aliquot was achieved.

Metagenomic functional screening of hexosaminidase active clones

Approximately 100,000 colony forming units from the metagenomic library were plated on LB agar plates supplied with 2 mL/L fosmid induction solution, 1 μg/mL X-GlcNAc, 34 μg/mL chloramphenicol, and 100 μg/mL ampicillin. β-*N*-acetylhexosaminidase active clones were identified based on their ability to hydrolyze X-GlcNAc, yielding blue colonies after incubation at 30 °C for 24 to 48 h. It has been reported that *E. coli* can catabolize the X-GlcNAc analogue $GlcNAc_2$ (Keyhani et al. 2000). However, in our control experiments, we did not observe color formation.

Fosmid isolation of hexosaminidase active clones

Fosmids were isolated from blue colonies as follows. Two milliters of LB medium supplied with 2 μL/mL fosmid induction solution, 34 μg/mL chloroamphenicol, and 100 μg/mL ampicillin were inoculated with a single colony and grown overnight while shaking at 30 °C. Cells were pelleted at 1900×*g* for 10 min and fosmids isolated with FosmidMAX™ DNA Purification Kit (Epicentre, USA). Restriction analysis of isolated fosmids was performed with *Pst*I and *Pvu*II to check for redundancy of positive clones.

Subcloning of fosmid genes from active clones

Isolated fosmids from active clones were digested for 3 h at 37 °C with one of the three restriction enzymes *Bam*HI, *Eco*RI, and *Pst*I. Digested fragments were ligated into pUC18 digested with the identical restriction enzyme in the presence of shrimp alkaline phosphatase (SAP). DH5α electrocompetent cells were transformed and plated on LB agar plates supplemented with 100 μg/mL ampicillin and 100 μg/mL X-GlcNAc. After overnight growth, colored colonies were picked and the pUC18 plasmids containing the β-*N*-acetylhexosaminidase coding gene isolated. Inserts from isolated plasmids were excised with *Bam*HI, *Eco*RI, or *Pst*I, gel-purified, and sequenced by IonTorrent PGM sequencing using 316 chip (DMAC, Lyngby, Denmark) and Sanger sequencing. Sequence reads were trimmed using clc_quality_trim script (CLC BIO, Aarhus, Denmark) and assembled using Ray Meta (Boisvert et al. 2012). Closest homologues were identified using BlastX (Altschul et al. 1990). Sequencing primers were designed to close gaps between contigs containing β-*N*-acetylhexosaminidase genes. Full-length genes were amplified by PCR using primers shown in Table S1. The PCR products were restricted and inserted in the plasmid pETM-10 using the restriction sites *Nco*I and *Bam*HI or *Nco*I and *Eco*RI.

Expression and purification of novel β-N-acetylhexosaminidase

E. coli BL21 (DE3), Tuner (DE3), C41 (DE3), and C43(DE3) were transformed with pETM-10 harboring the insert of interest and were grown overnight at 30 °C and 150 rpm in either (i) a lactose auto-induction medium (Studier 2005) or (ii) LB medium supplemented with 1 M IPTG when the cell culture reached an OD_{600}=0.6. Cells were pelleted by centrifugation, re-suspended in binding buffer (20 mM Na-phosphate buffer, 500 mM NaCl, 20 mM imidazole, pH 7), and lysed by sonication. Cell debris was removed by centrifugation for 20 min at 5000×*g*. For identification of optimal expression conditions, 10 μL supernatant from each sample was analyzed by SDS-PAGE (Fig. S2). The bulk supernatant was passed through a 0.22 μm filter and loaded on a 5 mL Ni^{2+}-sepharose HisTrap HP column (GE Healthcare, Uppsala, Sweden), primed with binding buffer using an ÄKTA purifier (GE Healthcare, Uppsala, Sweden). After loading the sample, unbound material was washed off the column with 10 column volumes (CV) of binding buffer. The β-*N*-acetylhexosaminidases were eluted with a linear gradient of 0-100 % elution buffer (20 mM Na-phosphate buffer, 500 mM NaCl, 500 mM imidazole, pH 7) in 10 CV. Protein purity was assessed by SDS-PAGE, the identity confirmed by MALDI-TOF (Fig. S3) and the concentration determined using absorption at 280 nm in combination with an

extinction coefficient of 87,500 M^{-1} cm^{-1} (HEX1) and 97, 400 M^{-1} cm^{-1} (HEX2), respectively (Wilkins et al. 1999).

Catalytic characterization of novel enzymes

The pH optima for the β-*N*-acetylhexosaminidases were determined in a discontinuous assay using 10 mM phosphate-citrate buffer (McIlvaine 1921) with pH values ranging from 2.7 to 10.0. Reactions were run at 25 °C with 5 mM *p*NP-GlcNAc as substrate and were stopped after 15 s and 1.5, 3, 4.5, and 6 min by addition of Na_2CO_3. Absorbance was measured at 410 nm in an Infinite200 microplate reader (Tecan, Austria).

Thermal stability was determined by incubation of the β-*N*-acetylhexosaminidases at 45, 50, 55, and 60 °C for up to 60 min prior to determination of remaining activity at standard conditions, i.e., 25 °C, 10 mM phosphate-citrate buffer at optimal pH (pH 8 for HEX1 and pH 6 for HEX2) and 5 mM *p*NP-GlcNAc. Reactions were conducted as described above.

Michaelis–Menten kinetics for enzyme-catalyzed hydrolysis of *p*NP-GlcNAc were established for both β-*N*-acetylhexosaminidases in a discontinuous assay at 25 °C in 10 mM phosphate-citrate buffer at several pH values (McIlvaine 1921). A stock solution of 15 mM *p*NP-GlcNAc was prepared and diluted to give the desired final substrate concentrations. Reactions were stopped after 1, 2, 3, and 4 min by addition of Na_2CO_3. Absorbance measurements and data collection were performed as described above.

Hydrolysis of *p*NP-GalNAc was performed in the presence of dimethylformamide (DMF) due to its low solubility in water and compared with pNP-GlcNAc hydrolysis under identical conditions.

Substrate screening

The substrate versatility of the enzymes was tested on colloidal chitin, chitosan, and the chitin oligomers *N*,*N*′-diacetylchitobiose, *N*,*N*′,*N*″-triacetylchitotriose and *N*,*N*′,*N*″,*N*‴-tetraacetylchitotetraose. Colloidal chitin was prepared according to Rojas-Avelizapa et al. (1999): 10 g chitin were dissolved in 100 mL 85 % (*v*/*v*) phosphoric acid, incubated overnight at 4 °C and washed with milli-Q water until complete acid removal. Residual water was removed with a Büchner funnel, and the colloidal chitin was used without further filtering. Chitosan was used in various concentrations by dilution from a 1 % (*w*/*v*) stock in 1 % (*v*/*v*) acetic acid. *N*-acetyl-chitooligosaccharides were assayed with a concentration of 30 mM at 25 °C in 10 mM Na-phosphate citrate buffer at pH 8 and 2.6 μM enzyme (HEX1) or pH 6 and 2.2 μM enzyme (HEX2). Reactions were stopped by diluting the reaction mix 20 times with 99 °C hot milli-Q water, and the enzyme was removed by filtration using Vivaspin 500 centrifugal concentrators with 5000 MWCO (Sartorius, Germany). The permeate was analyzed by HPAEC-PAD.

Trans-glycosylation reactions

Trans-glycosylation reactions were performed at 25 °C in the presence of 10 mM Na-phosphate citrate buffer at pH 8 (HEX1) or pH 6 (HEX2); 100 mM*N*,*N*′-diacetylchitobiose was used as GlcNAc donor substrate (D) and 500 mM β-lactose (or other carbohydrates) served as acceptor (A) with an enzyme to substrate ratio (*E*:*S*) of 0.24 % (*w*/*w*). After 1–2 h, reactions were stopped by diluting the reaction mix in 99 °C hot milli-Q water, followed by filtration using Vivaspin 500 columns (10,000 MWCO; Sartorius, Germany) thereby removing the enzyme. The permeate was analyzed by high-performance anion exchange chromatography (HPAEC-PAD), and trans-glycosylation products were identified by mass spectrometry. The yield of LNT2 and an isomeric form of LNT2 formed by HEX1 and HEX2 was quantified after 2 h reaction time by HPAEC-PAD, integrating the peak at the retention time where the external LNT2 standard eluted and calibrating it against a standard curve based on the external LNT2 standard. The yield of the isomeric form of LNT2 was quantified by integrating the peak eluting after the LNT2 peak and calibrating it against the LNT2 external standard curve, assuming an equal PAD signal for the LNT2 isomer.

High-performance anion exchange chromatography

Separation and quantification of products from biocatalytic reactions were carried out using a Dionex BioLC system consisting of GS50 gradient pumps, an ED50 electrochemical detector, and an AS50 chromatography compartment coupled to an AS50 autosampler (Dionex Corp., Sunnyvale, CA). Samples of 10 μL were injected on a CarboPac™ PA1 (4× 250 mm) analytical column equipped with a CarboPac™ PA1 (4×50 mm) guard column (Dionex Corp., Sunnyvale, CA). Products from trans-glycosylation and hydrolysis reactions were eluted under isocratic conditions with 75 mM NaOH in deionized water for 35 min at a flow rate of 1 mL/min. Strongly retained anions were subsequently eluted from the column by 75 mM NaOH including a gradient of NaOAc supplemented with 0.02 % (*w*/*v*) NaN_3 from 0 to 100 mM in 5 min. The column was washed with 425 mM NaOAc/75 mM NaOH for 5 min followed by re-equilibration at starting conditions for 5 min.

Mass spectrometry

MALDI-TOF of the purified β-*N*-acetylhexosaminidases was performed on an Ultraflex II MALDI-TOF/TOF (Bruker, USA). Protein samples were desalted using Vivaspin 500

centrifugal concentrators with a MWCO of 5000 Da (Sartorius, Germany) and mixed 1:1 (*v*/*v*) with a 10 % (*w*/*v*) sinapinic acid solution dissolved in 70 % acetonitrile and 0.1 % trifluoroacetic acid, before being spotted on a MTP 384 ground steel target plate (Bruker, USA).

The enzyme-catalyzed formation of LNT2 was analyzed by MALDI-TOF and MALDI-TOF/TOF and compared with the commercial standard. A 10-mg/mL norharmane solution was prepared in a 1:1 mixture of acetonitrile and water. After spotting 1-μL sample solution to a MTP 384 ground steel target plate (Bruker, USA), 1-μL matrix was added and the sample plate dried completely prior measurement.

Trans-glycosylation products were analyzed by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry (HILIC-ESI/MS^n) in the negative mode. An Accela UHPLC system (Thermo Scientific, Waltham, MA, USA) was coupled to a Velos Pro mass spectrometer (LTQ Velos Pro ion trap MS, Thermo Scientific). The chromatographic separation was performed on an Acquity UPLC BEH Amide column (1.7 μm, 2.1×150 mm) combined with a Van Guard precolumn (1.7 μm, 2.1×5 mm; Waters Corporation, Milford, MA, USA). The samples were dissolved in ACN/H_2O 50:50 (*v*/*v*), and 5 μL were injected. The elution flow rate was 300 μL/min with a composition of the mobile phases of: (A) water with 1 % (*v*/*v*) ACN, (B) 100 % (*v*/*v*) ACN, and (C) 200 mM ammonium formate buffer (pH 4.5). The elution was performed as follows: 1 min isocratic 85 % B; 30 min linear gradient from 85 to 60 % B; 5 min linear gradient from 60 to 40 % B; followed by 9 min of column washing with a linear gradient from 40 to 85 % B. The eluent C was kept constant at 5 % during the elution program.

Protein homology modeling

Protein homology models were generated with the CPHmodels Server 3.2 (Nielsen et al. 2010) on the basis of 1HP4 (homologue to HEX1) and 3RCN (homologue to HEX2), respectively.

Sequence alignment and phylogenetic tree

Amino acid sequence alignments and rooted phylogenetic trees were made with Phylogeny.fr (Dereeper et al. 2008) and were visualized with FigTree, downloaded from http://tree.bio.ed.ac.uk/software/figtree/.

Data analysis

Non-linear regression analysis was performed using ProFit (Quantumsoft, Switzerland) and COMSOL Multiphysics 4.4 (COMSOL, Inc., USA).

Accession numbers

The complete nucleotide sequences of the novel β-*N*-acetylhexosaminidase encoding genes are available in GenBank with the submission numbers: HEX1: KP893200 (called SoMe-cHEX1) and HEX2: KP893201 (called SoMe-cHEX2).

Results

Screening, subcloning, and identification of novel β-*N*-acetylhexosaminidases

To identify novel hexosaminidases, a fosmid-based soil metagenomic library containing approximately 100,000 colony forming units was screened on LB agar plates supplied with the chromogenic substrate 5-bromo-4-chloro-3-indolyl-*N*-acetyl-beta-D-glucosaminide (X-GlcNAc). Color formation due to β-*N*-acetylhexosaminidase activity was observed for 30 clones. From the 30 clones, non-redundant fosmids were digested with *Bam*HI or *Pst*I and ligated into pUC18. *E. coli* transformed with the resulting construct was screened on plates supplemented with X-GlcNAc resulting in approximately 1 % blue colonies. Plasmids from positive clones were isolated and the inserts sequenced (annotated sequence contigs shown in Fig. S1). The open reading frame was identified for the genes encoding β-*N*-acetylhexosaminidases by identification of the start and stop codons as well as of the putative ribosomal binding sites. The genes encoding the novel β-*N*-acetylhexosaminidases were denoted HEX1 and HEX2.

Description of novel β-*N*-acetylhexosaminidases

The genes encoding the two identified β-*N*-acetylhexosaminidases HEX1 and HEX2 have a length of 1461 and 1614 nucleotides, corresponding to a protein length of 494 amino acids for HEX1 and 537 amino acids for HEX2. The genes encoding HEX1 and HEX2 share protein identities of 51 and 59 % with β-*N*-acetylhexosaminidases from *Streptomyces scabiei* (HEX1) and *Algibacter lectus* (HEX2), their respective closest homologues. They were classified by Blastp (Altschul et al. 1990) and CaZYmes (Park et al. 2010) to be members of glycoside hydrolase family 20 (GH 20). A sequence alignment of the catalytic domains of HEX1 and HEX2 and their respective closest homologues is shown in Fig. 1. A phylogenetic tree of HEX1, HEX2, and closely related β-*N*-acetylhexosaminidases, as well as the β-*N*-acetylhexosaminidases from *Canavalia ensiformis* (jack bean) and *A. oryzae* used in the literature is shown in Fig. 2. HEX1 clusters in a clade with β-*N*-acetylhexosaminidases from organisms in the order of *Actinomycetales*, while HEX2 clusters in a clade with β-*N*-acetylhexosaminidases from the phylum

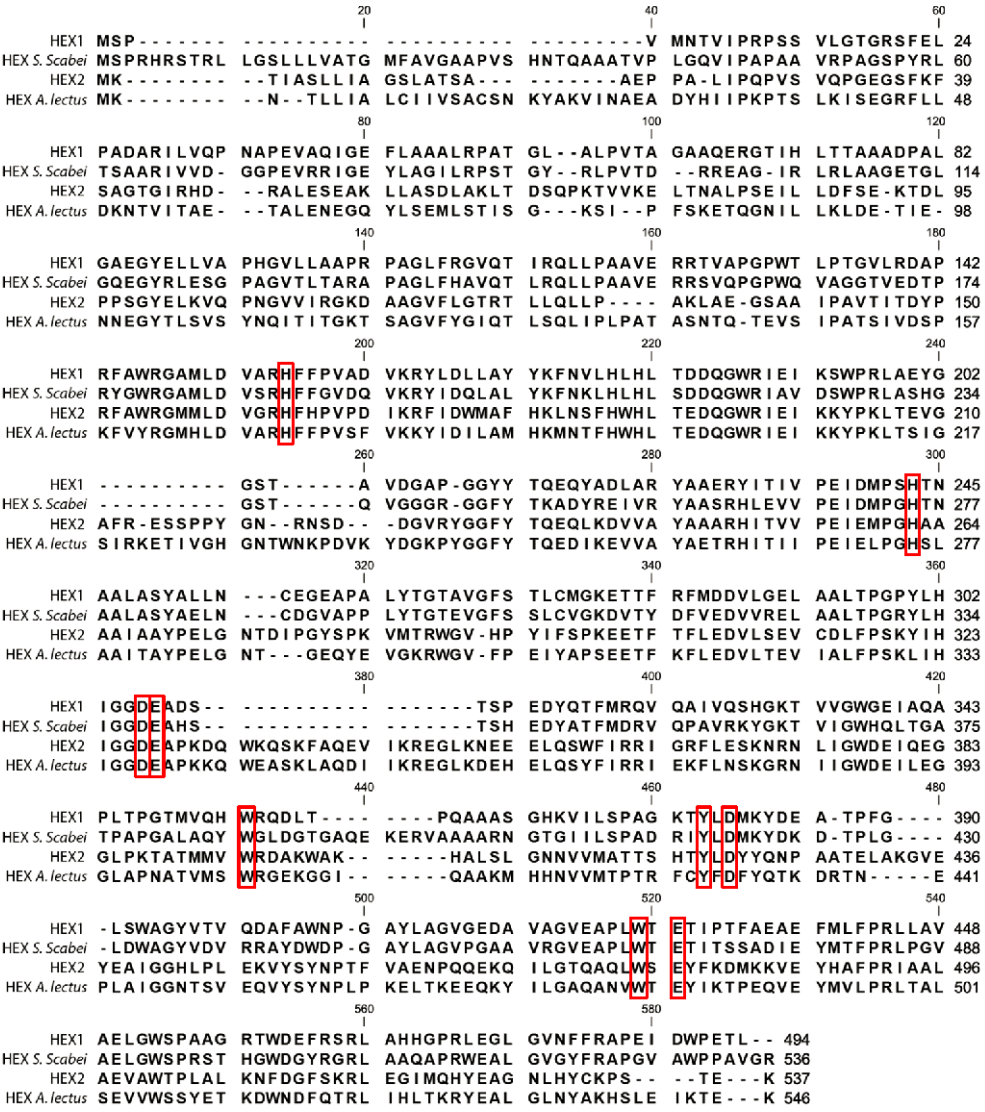


Fig. 1 Amino acid sequence alignment of β-*N*-acetylhexosaminidases HEX1 and HEX2 and their respective closest homologues (in *line* below, verified β-*N*-acetylhexosaminidases). Conserved active site residues are highlighted (*red boxes*)

Bacterioidetes. Characteristics of HEX1 and HEX2 are summarized in Table 1, and predicted structures were aligned with their respective closest homologue, for which a structure was available (Fig. 3).

Expression and purification novel β-*N*-acetylhexosaminidases

IPTG-induced enzyme expression in BL21 (DE3) was insufficient for both enzymes but could be improved significantly using the strains C41 (DE3; for expression of HEX1) and C43 (DE3; for expression of HEX2), respectively. Further improvements were achieved by using an auto-induction medium (Studier 2005; Görke and Stülke 2008) (Fig. S2), resulting in 20 mg purified enzyme from 2.4 L culture medium. Both enzymes could be purified by nickel affinity chromatography and their target masses verified by MALDI-TOF mass spectrometry (Fig. S3).

Catalytic characterization of novel enzymes

HEX1 exhibits a local optimum at pH 4.5 and a global optimum around pH 8, while HEX2 exhibits a single pH optimum around pH 6 (Fig. S4.

At 45 and 50 °C, the thermal inactivation constant k_D is virtually zero for both enzymes, indicating that the β-*N*-acetylhexosaminidases are stable at these temperatures for at least 60 min (Fig. 4). At the higher temperatures tested, k_D increased from approximately 0.1 min^{-1} at 55 °C to 0.3 min^{-1} at 60 °C, resulting in a half-life time around 10 and 3 min, respectively. The Michaelis–Menten parameters for the two β-*N*-acetyl-hexosaminidases are summarized in Table S2. Additionally to *p*NP-GlcNAc, HEX1 and HEX2 are able to

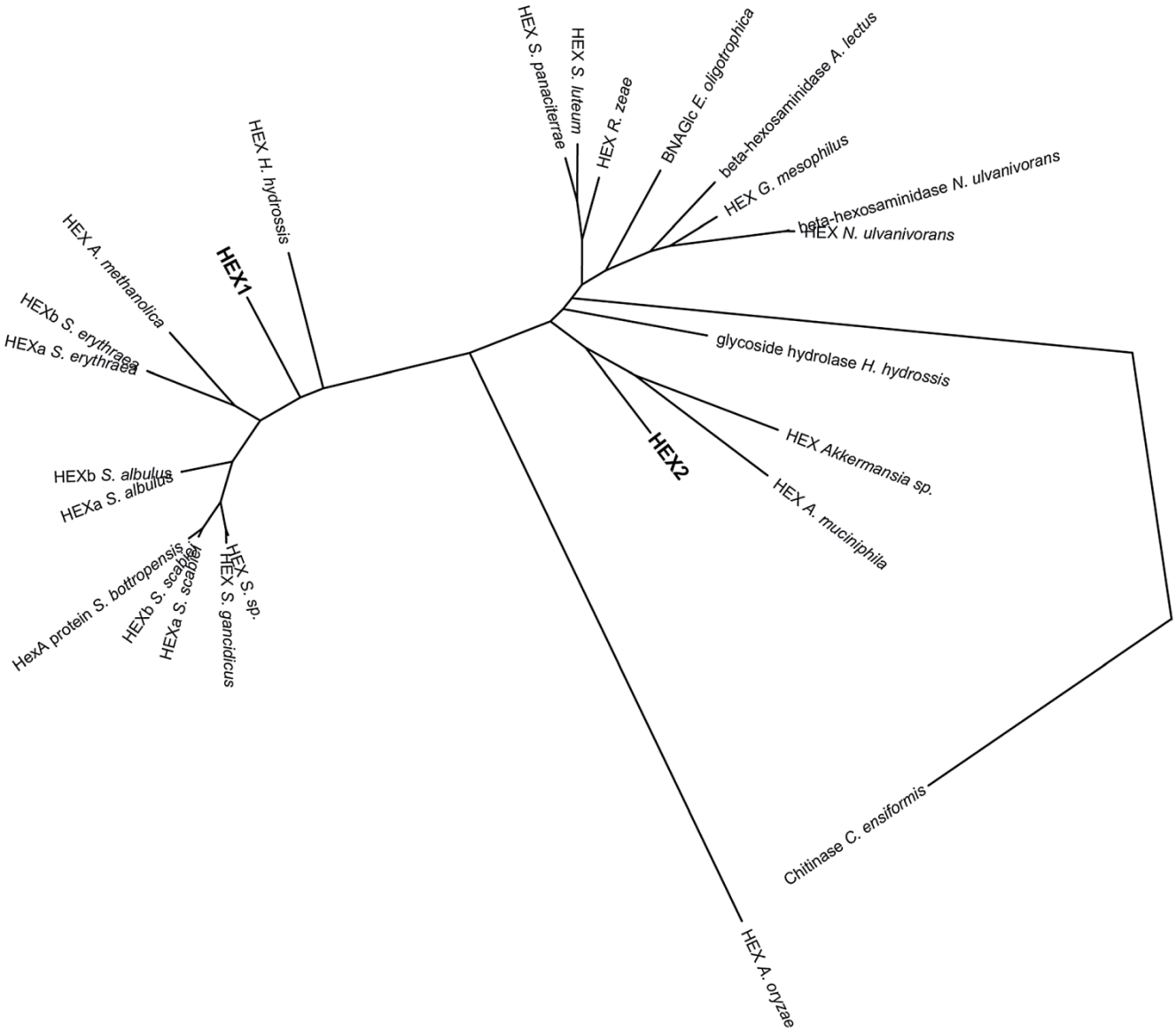


Fig. 2 Rooted phylogenetic tree of HEX1 and HEX2, their closest homologues, and β-*N*-acetyl hexosaminidases used in literature. *HEX:* β-*N*-acetylhexosaminidase, *BNAGlc:* β-*N*-acetylglucosaminidase. Enzymes in the phylogenetic tree originate from *Spirosoma panaciterrae*, *Spirosoma luteum*, *Runella zeae*, *Emticicia oligotrophica*, *Algibacter lectus*, *Gelidibacter mesophilus*, *Nonlabens ulvanivorans*, *Canvalia ensiformis*, *Haliscomenobacter hydrossis*, *Akkermansia muciniphila*, *Aspergillus oryzae*, *Streptomyces gancidicus*, *Streptomyces scabiei*, *Streptomyces bottropensis*, *Streptomyces albulus*, *Saccharopolyspora erythraea*, *Amycolatopsis methanolica*, *Haliscomenobacter hydrossis*

hydrolyze *p*NP-GalNAc, although with a lower activity for the latter substrate.

Utilizing the novel β-*N*-acetylhexosaminidases for trans-glycosylation reactions

Screening of substrates to assess their potential as donors in trans-glycosylation reactions

In addition to the non-natural substrate *p*NP-GlcNAc, we tested the hydrolytic activity of the two novel β-*N*-acetylhexosaminidases for a series of carbohydrates containing the building block GlcNAc. None of the enzymes are able to degrade colloidal chitin or chitosan but both show activity towards *N*-acetyl-chitooligosaccharides. Both enzymes are capable of hydrolyzing *N*,*N*′-diacetylchitobiose ($GlcNAc_2$), *N*,*N*′,*N*″-triacetylchitotriose ($GlcNAc_3$), and *N*,*N*′,*N*″,*N*‴-tetraacetylchitotetraose ($GlcNAc_4$) resulting in the formation of GlcNAc. In the case of the latter two substrates, the respective intermediates, shortened by one or two GlcNAc residues could be observed.

Table 1 Characteristics of identified β-*N*-acetylhexosaminidases on the protein level

Name	Protein length (amino acids)	Molecular weight (kDa)	Protein identity (similarity; %)	Best protein hit (accession number)	Organism name of best protein hit
HEX1	494	53.4	59 (70)	β-*N*-acetyl hexosaminidase (KFG03330.1)	*Streptomyces scabiei*
HEX2	537	60.2	51 (64)	β-*N*-acetyl hexosaminidase (GAL64280.1)	*Algibacter lectus*

Trans-glycosylation reactions and synthesis of LNT2

Both enzymes were used in biocatalytic reactions in order to form the HMO backbone precursor LNT2 by trans-glycosylation from $GlcNAc_2$ to β-lactose (Fig. 5). Two prominent trans-glycosylation products could be observed after one hour reaction by HILIC-ESI/MS^n. The first peak eluting after 18.1 min was identified as the product LNT2

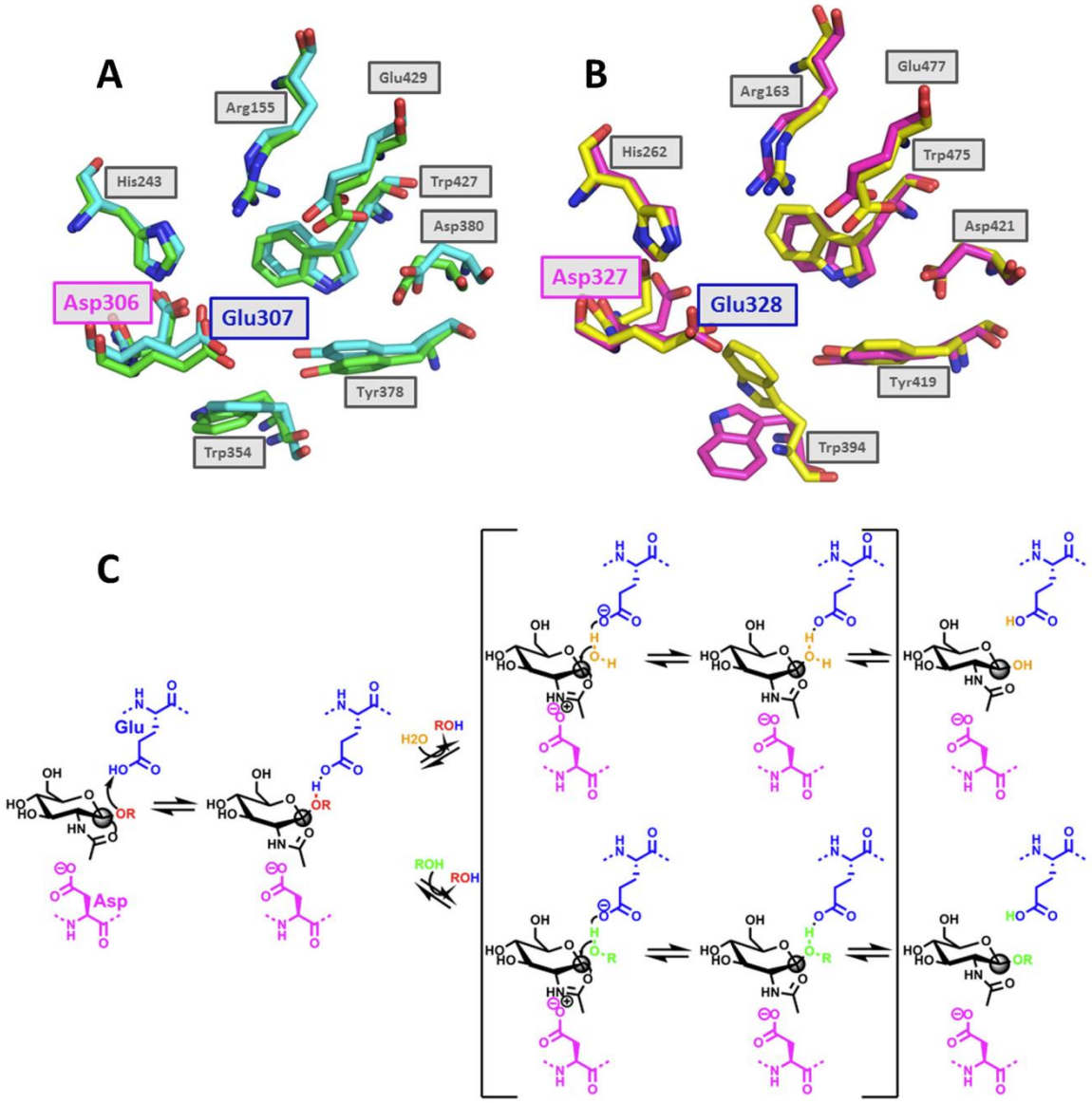


Fig. 3 Catalytic sites and mechanism of β-*N*-acetylhexosaminidases from GH 20. Overlay of catalytic site conserved amino acid residues from a HEX1 homology model and 1HP4 (**a**) (Mark et al. 2001) and a HEX2 homology model in combination with 3RCN (**b**) (Kim et al. 2011). **c** Double displacement retaining mechanism (Slámová et al. 2010) with the anomeric center *highlighted with a sphere*. The catalytic Glu serving as a general acid/base is shown in *blue*, and the neighboring Asp stabilizing the oxazoline intermediate is shown in *pink*. Hydrolysis of the substrate by water (*yellow*) is shown in the *upper path*; transfer of the GlcNAc moiety to the non-reducing end of a carbohydrate (*green*) is shown in the *lower path*. Intermediates are indicated by *brackets*

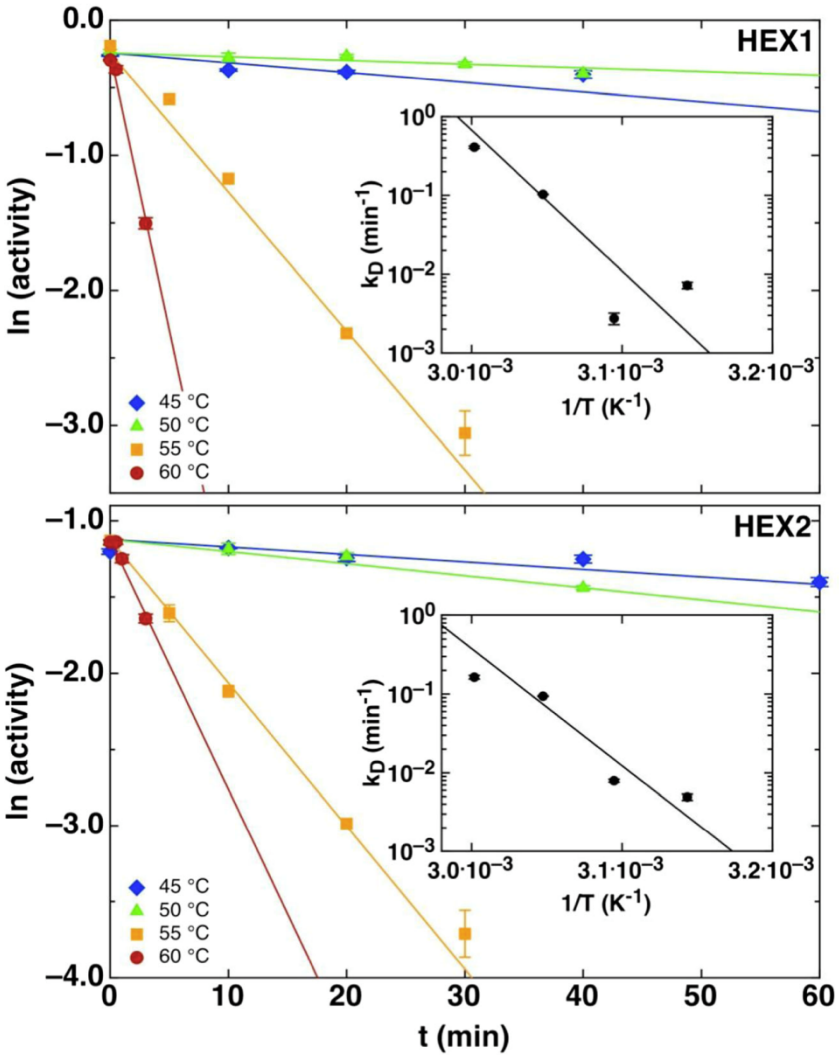


Fig. 4 Remaining *p*NP-GlcNAc hydrolytic activity after heat inactivation of HEX1 and HEX2. Enzymes were incubated for the indicated time at temperatures between 45 and 60 °C. Remaining activity was determined at 25 °C in 10 mM phosphate-citrate buffer at pH 8 (*HEX1*) or pH 6 (*HEX2*). The rate constants for thermal inactivation (k_D) are shown in the Arrhenius plots as a function of the inverse temperature (*inserts*)

with identical retention time and MS_2 spectrum as the LNT2 standard. A second peak eluting after 18.9 min had the identical total mass as LNT2, yet showed a different MS_2 spectrum when compared with the LNT2 standard (see Fig. 6). From the fragmentation pattern shown in the MS^2-spectrum of peak 18.9 min, it is speculated that the LNT2 isomer produced by HEX2 could represent a *N*-acetyl-glucosamine-β-1,4-linked to the non-

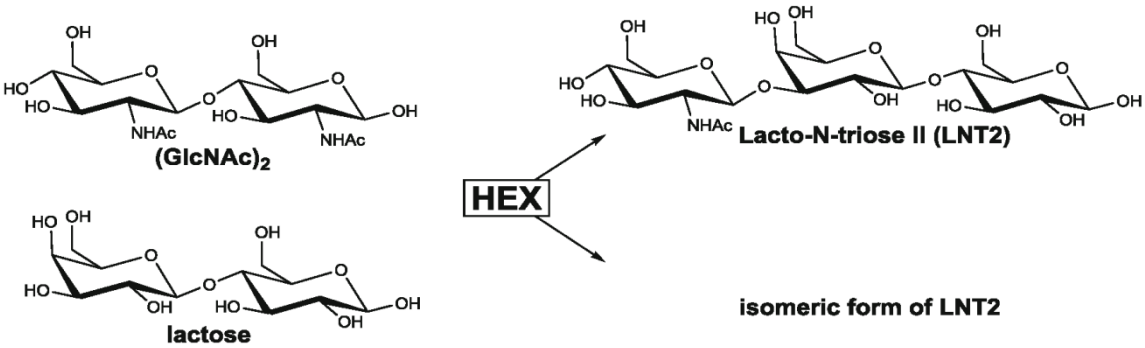


Fig. 5 Reaction scheme for β-*N*-acetylhexosaminidase (*HEX*) catalyzed production of LNT2 (1-3 linkage between GlcNAc and Gal) and an isomeric form of LNT2. Product formation was verified by HILIC-ESI/MS^n and MALDI-TOF/TOF (Table 2; Fig. 6; Fig. S5) and quantified by HPAEC-PAD (Table S3)

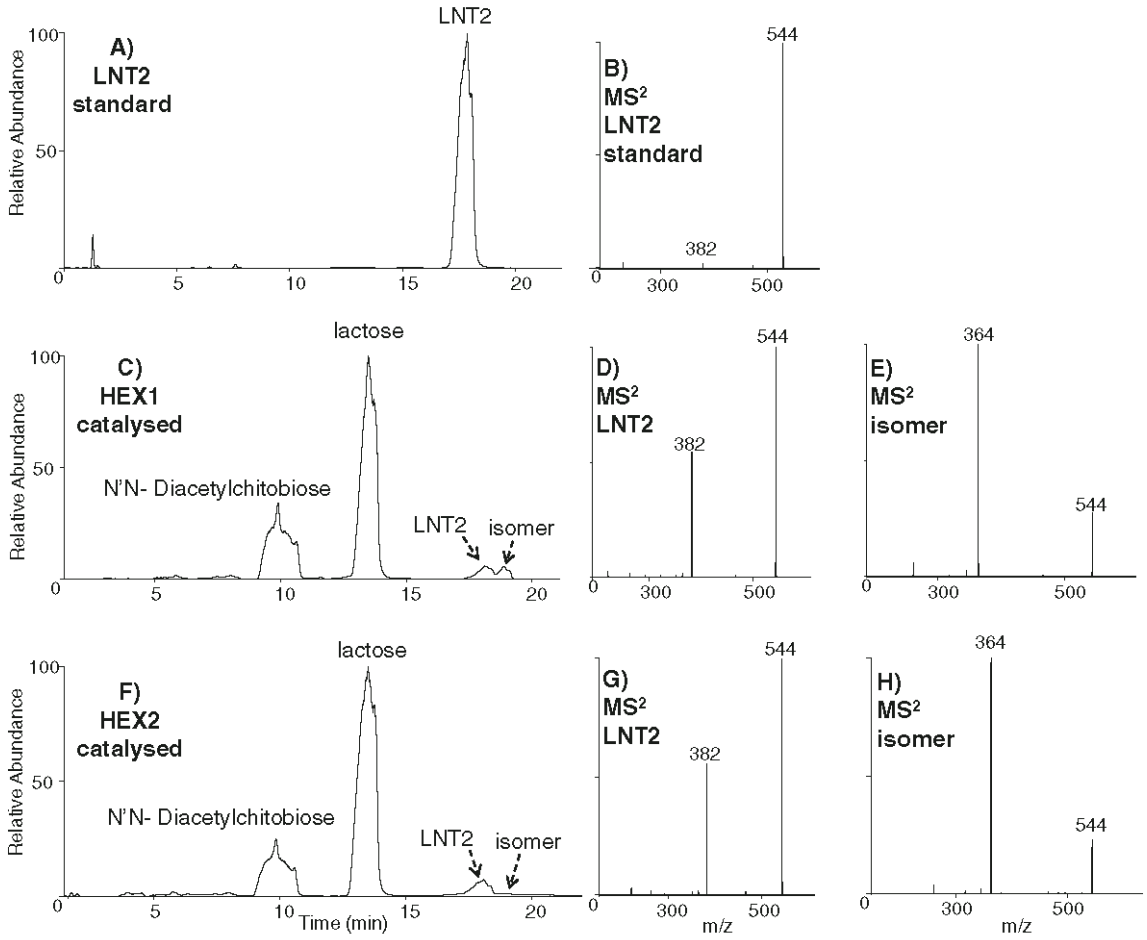


Fig. 6 HILIC-ESI/MS^n elution pattern of transglycosylation reactions of HEX1 (**c**) and HEX2 (**f**) with $GlcNAc_2$ as donor and β-lactose as acceptor. Transglycosylation reactions were performed for 60 min at 25 °C with 100 mM *N,N'*-diacetylchitobiose (donor) and 500 mM β-lactose (acceptor) and an enzyme to substrate ratio (*E*:*S*) of 0.24 % (*w*/*w*). Non-hydrolyzed donor *N,N'*-diacetylchitobiose and unreacted acceptor β-lactose are indicated. The desired product LNT2 elutes after 18.1 min with a mass of 544 Da (commercial standard shown in (**a**)). An additional peak can be observed for the enzyme-catalyzed reactions, due to the formation of an isomeric form of LNT2

reducing end of galactose. The anotation of LNT2 was further proven by MALDI-TOF (Fig. S5) and the LNT2 yield determined to be 2 % (HEX1) and 8 % (HEX2), respectively (Table S3).

In addition to β-lactose, trans-glycosylation reactions were performed with the monomeric acceptors galactose, glucose and the dimeric acceptors sucrose and maltose in combination with the donor $GlcNAc_2$. Both β-*N*-acetyl-hexosaminidases were able to attach GlcNAc to the donor compounds, i.e., the masses corresponding to the addition of a GlcNAc to the respective acceptor could be detected by HILIC-ESI/MS^n. Also, (i) masses corresponding to the addition of two GlcNAc residues to the acceptor were recorded, as well as (ii) masses corresponding to the elongation of the acceptor by another acceptor molecule including a GlcNAc residue. The masses recorded and the corresponding oligosaccharides are represented in Table 2. For several species, a formiate anion adduct was found.

Discussion

Characterization

In this paper, we report the identification and characterization of two novel β-*N*-acetyl-hexosaminidases denominated as HEX1 and HEX2 originating from a soil metagenomic library. The two identified β-*N*-acetylhexosaminidases presumably originate from soil bacteria, since they cluster with enzymes from soil bacteria in the phylogenetic tree (Fig. 2). HEX1 clusters with a clade of β-*N*-acetylhexosaminidases mainly from *Actinomycetales*, while HEX2 lies in the middle between the enzymes from *A. oryzae* and *C. ensiformis*, making this enzyme an interesting link between the different clades and branches. β-*N*-acetyl-hexosaminidases can be found in three glycoside hydrolase (GH) families, namely GH 3, GH 20, and GH 84 (Slámová et al. 2010). However, based on their sequence, HEX1 and HEX2 were classified as members of

Table 2 Trans-glycosylation products formed using HEX1 or HEX2, analyzed by HILIC-ESI/MS^n and verified by MS^2

	Acceptor	*m/z*			Products
		$[M-H]^-$	$[M+HCOO^-]^-$	$[M+(HCOO^-)_2+H^+]^-$	
HEX1	β-lactose	544	590	–	GlcNAc-1,3-β-Gal-1,4-β-Glc (LNT2)[a]
		544	590	–	Isomeric form of LNT2
		747	793	–	GlcNAc-GlcNAc-Lac[b]
		868	914	–	GlcNAc-Lac-Lac[b]
	Maltose	544	590	–	GlcNAc-(Glc-1,4-α-Glc)[b]
	Sucrose	–	590	–	GlcNAc-(Glc-1,2-α-Fru)[b]
		–	793	–	$GlcNAc_2$-(Glc-1,2-α-Fru)[b]
		–	913	–	GlcNAc-$(Glc\text{-}1,2\text{-}\alpha\text{-}Fru)_2$[b]
	Galactose	382	–	–	GlcNAc-Gal[b]
	Glucose	–	428	–	GlcNAc-Glc[b]
		–	631	–	$GlcNAc_2$-Glc[b]
		–	590	–	GlcNAc-Glc_2[b]
HEX2	β-lactose	544	590	–	GlcNAc-1,3-β-Gal-1,4-β-Glc (LNT2)[a]
		544	590	–	Isomeric form of LNT2
		747	793	–	GlcNAc-GlcNAc-Lac[b]
		868	914	–	GlcNAc-Lac-Lac[b]
	Maltose	544	590	–	GlcNAc-(Glc-1,4-α-Glc)[b]
	Sucrose	–	590	–	GlcNAc-(Glc-1,2-α-Fru)[b]
		–	793	–	$GlcNAc_2$-(Glc-1,2-α-Fru)[b]
		–	913	–	GlcNAc-$(Glc\text{-}1,2\text{-}\alpha\text{-}Fru)_2$[b]
	Galactose	382	–	–	GlcNAc-Gal[b]
	Glucose	–	428	–	GlcNAc-Glc[b]
		–	631	–	$GlcNAc_2$-Glc[b]
		–	590	–	GlcNAc-Glc_2[b]

[a] Linkage determined by comparing MS and MS/MS data with LNT2 standard

[b] Linkage not determined

Abbreviations: *LNT2* lacto-N-triose II, *GlcNAc* β-*N*-acetylglucosamine, *Gal* galactose, *Glc* glucose, *Lac* β-lactose, *Fru* fructose

GH 20 (Altschul et al. 1990). Enzymes in GH 20-fold into a TIM barrel, containing eight repeats of a helix-strand motif and contain nine conserved residues forming the catalytic site (Arg, Asp_2, Glu_2, His, Trp_2, Tyr, see Fig. 1) (Kim et al. 2011; Mark et al. 2001). The catalytic Glu residue serves as proton donor during hydrolytic catalysis, attacking the nucleophile carbonyl oxygen of the C-2-acetamido group of the substrate (Park et al. 2010; Lombard et al. 2014). In the enzymes described here, all conserved residues are present and the catalytic Glu could be located at positions 307 (HEX1) and 328 (HEX2) according to the multiple alignment (Fig. 1) and 3D structure comparison (Fig. 3a, b). The conserved hydrophobic residues are expected to be associated with carbohydrate binding and might influence the accessibility of water molecules to the catalytic site (Amaya et al. 2004; Jers et al. 2014).

With their size of 53.4 kDa (HEX1) and 60.2 kDa (HEX2), the identified enzymes are amongst the smallest bacterial β-*N*-acetylhexosaminidases deposited in the Braunschweig Enzyme Database (BRENDA) to date (Schomburg et al. 2000).

In order for enzymes to be indeed classified as β-*N*-acetylhexosaminidases, they must have both β-D-*N*-acetylgalactosaminidase and β-D-*N*-acetylglucosaminidase activity (Cabezas 1989). The identified enzymes HEX1 and HEX2 are thus true β-*N*-acetylhexosaminidases, verified not only by sequence but also by activity. Both enzymes show significant activity over a broad pH range, i.e., between pH 3 and pH 9 for HEX1 (pH 5 and pH9 for HEX2). pH optima of all β-*N*-acetyl-hexosaminidases available on BRENDA have values between pH 2 and pH 8, making HEX1 one of the β-*N*-acetylhexosaminidases with the highest pH optimum reported to date.

The enzymes HEX1 and HEX2 were unstable at elevated temperatures, i.e., above 50 °C. They originate from mesophiles and did not seem to evolve towards high thermostability, due to the lack of evolutionary pressure.

While the enzymes are unable to degrade chitin or chitosan, *N*-acetyl-chito-oligosaccharides of different length (DP 2–4, DP>4 not tested) are accepted as substrates by both enzymes.

This is a great advantage, since these oligosaccharides can be conveniently produced using chitinase-assisted hydrolysis of chitin (Synstad et al. 2008). Chitin hydrolysis reaction products do not need to be separated as they are all useful substrates for HEX1 and HEX2. This saves cost and time. Furthermore, this will increase trans-glycosylation yields due to the availability of more than just one GlcNAc moiety (available GlcNAc moieties=DP−1) per chito-oligosaccharide donor, which can be attached to the acceptor. The chitinase substrate chitin is the second most abundant polysaccharide in nature after cellulose (Tharanathan and Kittur 2003). It is relatively cheap and can be harnessed from industrial food waste streams (Shahidi et al. 1999).

Trans-glycosylation reactions

β-*N*-acetylhexosaminidases have been used in literature to produce the human milk oligosaccharide backbone structure LNT2 and variants thereof (Matahira et al. 1995; Murata et al. 1997; Matsuo et al. 2003). Matsuo et al. used a β-*N*-acetyl-hexosaminidase from *A. oryzae* to transfer free GlcNAc to β-lactose and were able to synthesize LNT2 with a low yield of 0.2 % (Matsuo et al. 2003). Higher yields of 0.5 % (based on the amount of $GlcNAc_2$ added) were obtained by Murata et al. in a comparable reaction setup, where a β-*N*-acetylhexosaminidase from *Nocardia orientalis* was used to transfer GlcNAc from $GlcNAc_2$ to *p*-nitrophenyl β-*N*-acetyllactosaminide (Murata et al. 1997). Although $GlcNAc_2$ is cheap and can be conveniently obtained from chitin, the cost effectiveness of this reaction is limited due to the expensive acceptor *p*-nitrophenyl β-*N*-acetyllactosaminide. The same enzyme was used by Matahira et al. to synthesize methylated LNT2 from $GlcNAc_2$ and methyl-β-lactoside with a yield of 3.4 % based on the donor added (Matahira et al. 1995). The use of methyl-β-lactoside instead of β-lactose increases the cost of the overall reaction and results in the formation of a non-natural LNT2 variant. In this paper, we demonstrate the production of β-*N*-acetylhexosaminidase-catalyzed LNT2 formation from the cheap GlcNAc donor $GlcNAc_2$ and β-lactose (Figs. 5 and 6) with yields of 2 and 8 % based on the donor added (Table S3). Both HEX1 and HEX2 show good stereo-selectivity, making these enzyme interesting candidates for large-scale LNT2 synthesis. These findings highlight the options for biocatalytic synthesis of genuine human milk oligosaccharide backbone structures by transglycosylation technology using novel glycoside hydrolases identified through functional mining of metagenomic libraries.

Both β-*N*-acetylhexosaminidases were capable of attaching GlcNAc not only to β-lactose but also to monomeric and additional dimeric saccharide acceptors in trans-glycosylation reactions. Using the acceptor galactose, the concept of β-*N*-acetylhexosaminidase-assisted elongation of HMO backbones by addition of GlcNAc from di-*N*-acetylchitobiose to the non-reducing end of galactose was proven (Table 2). This elongation of the saccharide acceptor by a galactose is a necessary step in the elongation of HMO molecules (Kunz et al. 2000; Boehm and Stahl 2007; Bode 2012). The yield of trans-glycosylation products using acceptors different from β-lactose appears to be lower when compared with β-lactose. By trend, dimeric acceptors seem to give slightly higher trans-glycosylation yields. Nevertheless, the broad acceptor tolerance emphasized the versatility of the enzymes in trans-glycosylation reactions.

In order to improve the trans-glycosylation yields of the newly identified enzymes, enzyme engineering and reaction optimization, e.g., by optimizing the enzyme-to-substrate ratio or by the addition of co-solutes or co-solvents, are possible approaches.

Acknowledgments This work was supported by the Strategic Research Council in Denmark, http://fivu.dk; The project number is 09-067134, The title is "Enzymatic production of human milk oligosaccharides".

References

Altschul S, Gish W, Miller W (1990) Basic local alignment search tool. J Mol Biol 403–410

Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, Paris G, Frasch AC, Withers SG, Alzari PM (2004) Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure 12:775–84

Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni M-E, Polaina J, Strittmatter AW, Reva O, Waliczek A, Yakimov MM, Golyshina OV, Ferrer M, Golyshin PN (2010) Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl Environ Microbiol 76: 5934–46

Blixt O, van Die I, Norberg T, van den Eijnden DH (1999) High-level expression of the *Neisseria meningitidis* lgtA gene in *Escherichia coli* and characterization of the encoded N-acetylglucosaminyltransferase as a useful catalyst in the synthesis of GlcNAc beta 1→ 3Gal and GalNAc beta 1→3 Gal linkag. Glycobiology 9:1061–71

Bode L (2009) Human milk oligosaccharides: prebiotics and beyond. Nutr Rev 67(Suppl 2):S183–91

Bode L (2012) Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22:1147–62

Boehm G, Stahl B (2007) Oligosaccharides from milk. J Nutr 847–849

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J (2012) Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol 13:R122

Cabezas J (1989) Some comments on the type reference of the official nomenclature (IUB) for beta-N-acetylglucosaminidase, beta-N-acetylhexosaminidase and beta-N-acetylgalactosaminidase. Biochem J 261:1059–1061

Coppa G, Pierani P, Zampini L, Carloni I, Carlucci A, Gabrielli O (1999) Oligosaccharides in human milk during different phases of lactation. Acta Paediatr 430:89–94

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–9

Chapter 4: Enhancing trans-sialidase activity in a *Trypanosoma rangeli* sialidase by site-directed mutagenesis

4.1 Motivation

Genetically engineered mutants of the TrSA have been the preferred enzyme by our research group for trans-sialylation reactions. This is not least due to the fact that BioEng (or more accurately DTU) holds several patents on TrSA mutants. The patents will become commercially interesting if it is successfully established that the TrSA mutants are superior to other enzymes and that HMOs can be produced cost-efficiently. The primary motivation for this study was a publication by another research group who *in silico* predicted several novel mutations in the TrSA based engineered trans-sialidases which if implemented was claimed to improve trans-sialidase activity. Since no wet chemistry was carried out to confirm the predictions and since the base enzyme was different to Tr13 (the state of the art engineered trans-sialidase from this group), the motivation for this study was straight forward; to once more take the lead on engineering the TrSA into a better trans-sialidase. However the secondary motivation was clear as well; through implementation of the suggested mutations, establish better understanding of the reaction mechanism(s) of trans-sialylation by trypanosomal (trans-)sialidases.

4.2 Hypotheses and objectives

Three hypotheses from this study are relevant for the overall PhD project:

Hypothesis 4.1: The mutations suggested by Pierdominici-Sottile et al.(Pierdominici-Sottile et al., 2014) will improve trans-sialidase activity in TrSA mutants, but the trans-sialidase activity will be further improved by combination with the loop mutations suggested by Jers et al.(Jers et al., 2014).

Hypothesis 4.2: The improved trans-sialidase activity will enable efficient trans-sialylation at reaction conditions with low levels of acceptor substrate.

Hypothesis 4.3: By implementation of mutations in groups it will be possible to gain knowledge about the mechanistic traits on the individual mutations (in groups) contribution to trans-sialidase activity.

4.3 Experimental considerations

The main consideration to do with experimental setup in this study was how to implement the mutations to gain most possible knowledge about how the individual mutations contribute to trans-sialidase activity. It was decided to use a mutant first produced by Paris et al.(Paris et al., 2005) as a reference enzyme since it is claimed that the 5 mutations will be crucial to any TrSA trans-sialidase. That still left an additional 12 mutations and a complete combinatory mutagenesis strategy was therefore ruled out due to the workload that would be involved. Instead the mutations suggested by Jers et al.(Jers et al., 2014) was considered as one group whereas the mutations suggested by Pierdominici-Sottile et al.(Pierdominici-Sottile et al., 2014) was split into two groups because they had been predicted based on two different approaches. (Pierdominici-Sottile et al., 2014)

Another experimental consideration was dealt with during the study since one mutation led to weak or no expression. I would maybe have been possible to produce enough enzyme to carry out the reactions

required for comparison, if high density fermentations were carried out, but since such weak expressing enzymes would be of even weaker commercial interest, it was decided to disregard this mutations and remove it from the experimental design.

4.4 Conclusions

Two different mutants ($TrSA16_{mut}$ and $TrSA15_{mut}$) with dramatically improved trans-sialidase activity were produced by introduction of combinations of the suggested mutations. $TrSA16_{mut}$ was the enzyme with the lowest hydrolytic activity and was also the most mutated specimen with 16 of the suggested 17 mutations. The remaining mutation resulted in un-successful expression in all mutants into which it was introduced and a suggestion for causation was given. $TrSA15_{mut}$ was lacking one additional mutation, but whereas the hydrolytic activity was only slightly higher the overall activity of the enzyme was almost 20 X that of $TrSA16_{mut}$.

From the time series experiment using the popular donor substrate CGMP it was evident that the improved enzymes could efficiently produce 3'SL at a low acceptor concentration (with a donor to acceptor ratio of 1:4 compared to 1:88(Jers et al., 2014)).

By analyzing the combinatory introduction of enzymes it was possible to discuss the mechanism with which the groups of mutants affect the trans-sialidase activity of engineered TrSA-type mutants. However no definitive conclusions could be drawn in this regard and more studies will be required to unveil the remaining unknowns about the mechanism of the TcTS (and other trans-sialidases).

4.5 Paper III (Manuscript)

Improving trans-sialidase activity in a *Trypanosoma rangeli* sialidase by site-directed mutagenesis

Christian Nyffenegger, Rune Thorbjørn Nordvang, Carsten Jers, Anne S. Meyer, Jørn Dalgaard Mikkelsen*

Manuscript for submission to PLOS ONE

Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, Søltofts Plads, 2800 Kgs. Lyngby, Denmark.

*Corresponding author. Address as above. Phone: +45 4525 2909. Email: jdm@kt.dtu.dk.

Abstract

Two improved (superior) trans-sialidases from *Trypanosoma rangeli* with 15 and 16 mutations have been developed by rational design and free energy computations. One engineered trans-sialidases, $TrSA_{15mut}$, has a low hydrolysis rate and a higher rate of synthesis than any engineered trans-sialidase so far. The other, Tr_{16mut}, has the lowest hydrolytic activity of any engineered trans-sialidase so far but has a slower rate of synthesis. The wild type enzyme (TrSA) backbone is a true sialidase, but by several approaches whichin the past decade have sought (with varying success) to confer trans-sialidase activity to the enzyme. Five specific amino acids in the active site (M95V, A97P, S119Y, G248Y, Q283P) are believed to be crucial to trans-sialidase activity, while an additional 12 mutations (I36L, T38A, F58N, I196V, A197T, D198N, M199K, G200K, G201K, R202Q, D284G, G341A) have been suggested in various studies. In the present work, however, 7 novel mutants of the *T. rangeli* sialidase, having 6-16 targeted amino acid mutations, were constructed to enhance trans-sialidase activity of the *T. rangeli* sialidase for enzymatic production of human milk oligosaccharide (HMO) structures. The mutants were designed *in silico* from structural and energetic concepts based on the native *T. cruzi* trans-sialidase. Introduction of 15 and 16 mutations, respectively, produced significantly improved trans-sialylation activity of the *T. rangeli* enzyme for lactose sialylation. A total of 17 mutations were intended for this study, but one of the mutations (G341A) consistently led to insufficient expression. Combinatory introduction of mutations gave insight to the mechanisms whit which the mutations influenced trans-sialidase activity and surprisingly the findings were not completely consistent with the predictions.

Keywords: Trans-sialidase, TcTS, TrSA, Engineered sialidase, Human milk oligosaccharides, HMO

1. Introduction

Breast-fed neonates obtain high quantities of human milk oligosaccharides (HMO) from their mothers. These compounds (present in human milk in concentrations of 12-14 g/L[3]) are thought to improve the health of the nursling by (i) acting as prebiotics, (ii) preventing pathogen adhesion, (iii) modulation of immune and cell responses, and (iv) supplying nutrients for brain development [3]. The sialylated HMO, 3'-sialyllactose (3'SL), has been shown to reduce adhesion and invasion of Escherichia coli *in vitro*[3–6]. Moreover, 6'-sialyl-lactose was recently shown to reduce food allergy symptoms in vivo (in mice)[7].

Infant formulas are presently prepared with commercially available oligosaccharides such as FOS or GOS[8]. While the bifidogenic FOS and GOS may promote a healthy gut microbiota, imitating that of a breast-fed neonate, they are unlikely to replicate the multiple beneficial effects of the structurally diverse set of HMOs. Supplementing infant formula with sialylated oligosaccharides is believed to restore some of the beneficial effects of the HMOs and therefore their chemical or enzymatic synthesis is required. The latter strategy has been applied using sialyltransferases or trans-sialidases[1,9,10]. Using trans-sialidases for *in vitro* production of HMOs is more advantageous than transferases, since the former class of enzymes does not require expensive nucleotide-activated sialyl-donors but can utilize sialic acid (SA) from cheap donors[11]. A trans-sialidase from *Trypanosoma cruzi* (TcTS) exhibits high trans-glycosylation but low hydrolytic activity and has previously been used for the enzymatic elongation of a saccharide acceptor by a SA moiety[10]. *T. cruzi* is a human pathogen causing Chagas' disease where TcTS is an important virulence factor[12] and the TcTS is thus not suited for the production of food-grade HMOs. A closely related sialidase (TrSA) with 70% amino acid sequence identity exists in the non-pathogenic *Trypanosoma rangeli*, but lacks trans-sialidase activity[13]. Despite the high sequence identity between the TcTS and the

61 TrSA, identification of the amino acids necessary for trans-sialidase activity has proven to be
62 difficult [9,14]. Paris et al.[14] showed that five mutations M95V, A97P, S119Y, G248Y, and Q283P are
63 needed to confer detectable trans-sialidase activity but trans-sialidase activity could be further
64 improved by addition of either of the two mutations, I36L or G342A[14] (Since sequence numbering
65 in this article is based on TcTS, the numbering differs from the original literature). The mutations
66 I36L and G342A might affect the conformation and flexibility of the catalytic tyrosine-
67 nucleophile[14]. Trans-sialidase activities were 0.9% and 11.3% for the quintuple and sixtuple
68 mutants ($TrSA_{5mut}$ and $TrSA_{6mut}$), respectively when compared to the TcTS[14]. The sixtuple mutant,
69 $TrSA_{6mut}$ ($TrSA_{5mut}$/I36L) was expressed in *Pichia pastoris* and used for synthesis of 3'SL and other
70 sialylated glycans[14]. $TrSA_{6mut}$ has been further improved by mutation of a seven aminoacid motif
71 corresponding to TcTS amino acids 196–202 located in a loop at the border of the substrate
72 binding cleft of TrSA[3]. In the improved mutant ($TrSA_{13mut}$) the stretch is replaced with the
73 corresponding stretch from the TcTS which is strongly charged[1]. Although the mutated loop is 14 Å
74 away from the acceptor binding site, it reduces the hydrolytic activity without significantly
75 affecting the trans-sialidase activity – thereby allowing higher product yields (apparently due to
76 reduced product de-sialylation).
77 Using an *in silico* approach, the mechanisms responsible for the trans-sialidase activity of the TcTS
78 was recently studied by Pierdominici-Sottile and co-workers[2]. Free energy profiles for the
79 conversion of the Michaelis complex (MC) into the covalent enzyme-sialic acid intermediate (CI)
80 was computed for TcTS, TrSA, and $TrSA_{5mut}$, demonstrating that the covalent intermediate (CI) of
81 the TrSA is more stable than that of the TcTS[2]. The consequence of a higher CI stability is thought
82 to be a higher barrier for the reverse reaction, *i.e.* the coupling of sialic acid to an acceptor[2]. An
83 energy decomposition analysis, calculated the individual contributions of the 5 most CI stabilizing

84 residues in the TcTS. Among the five most CI stabilizing residues three (E96, T312, and E357) were
85 identical, whereas twodiffered (F58 and D284) in TrSA5mut. Hence it was suggested that mutation
86 of the two residues, F58 and D284 (to the corresponding TcTS residues N58 and G284), would
87 destabilize the CI. However, for the residues E96, T312 and E357, a different calculated energetic
88 contribution to the CI in the two enzymes TrSA5mut and the TcTS indicate that their interaction
89 with the active site residues (and substrate) differ in the two. Thus it is suggested that trans-
90 sialidase activity can be improved by introduction of mutations leading to adjustment of the five
91 stabilizing residues in $TrSA5_{mut}$[2]. Four of the mutations in $TrSA_{5mut}$ (aim to) change the
92 configuration of E96 and T312, and it is believed that only small improvements can be achieved
93 through further adjustment of the configuration of these two residues[2]. However, it is believed
94 that destabilizing the CI through a conformational adjustment of E357, and consequently the
95 interaction between this residue and the catalytic residues, is possible by introduction of a single
96 point mutation (T38A)[2]. In the same study it was proposed that two mutations (I37L and G342A)
97 would increase the mobility and consequently the participation of the nucleophile at the transition
98 state, which in turn would improve trans-sialidase activity. Thus introduction of the mutations
99 I36L, T38A, F58N, D284G, and G341A in $TrSA_{5mut}$ are proposed to yield an enzyme with high trans-
100 sialidase activity. However this (prior to this study) is not confirmed experimentally by
101 Pierdominici-Sottile et al.
102 In this paper, the findings of the three studies conducted by Paris et al., Jers et al., and
103 Pierdominici-Sottile et al. were integrated in an effort to design a sialidase derived from *T. rangeli*
104 with improved trans-sialidase activity and/or reduced hydrolytic activity. The suggested mutations
105 were split into three groups and systematic combination of these resulted in a total of seven

106 mutants which were constructed, expressed, and evaluated. This led to identification of two
107 mutants with improved trans-sialidase activity and reduced hydrolase activity.
108

109 2. Materials and methods

110 2.1 Chemicals

111 3'SLwas purchased from Carbosynth (Compton, United Kingdom). The commercial casein
112 glycomacropeptide (cGMP) product Lacprodan® CGMP-20 containing 5.7% (w/w) covalently linked
113 sialic acid was a gift from Arla Foods amba (Viby, Denmark). Before use, low molecular weight
114 impurities in the CGMP solution were removed by filtration using a 5 kDa membrane (Sartorius
115 AG, Goettingen, Germany) as a technical precaution relating to High-performance anion exchange
116 chromatography (HPAEC) analysis. All other chemicals were purchased from Sigma-Aldrich
117 (Steinheim, Germany).
118

119 2.2 DNA manipulations and strain construction

120 C-terminal His_6-tagged genes encoding (i) $TrSA_{5mut}$ + I36L and (ii) Tr13 for secreted expression in
121 *Pichia pastoris* were described earlier[1,9]. Mutations were introduced into these genes using the
122 QuikChange II Site-Directed Mutagenesis Kit (CA, USA) (Table 1). All plasmids were sequenced to
123 confirm the mutations and the integrity of the insert. Plasmids encoding the different enzyme
124 variants were propagated in *Escherichia coli* DH5α, cultured at 37 ºC while shaking in low salt LB
125 medium (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl), supplemented with 25 µg/mL zeocin.
126 *P. pastoris* X-33 was transformed with the *Mss*I-linearized vectors by electroporation following the
127 manufacturer's instructions (Invitrogen, CA, USA).

128

129 **2.3 Enzyme production and purification**

130 Shake-flask protein synthesis was performed by cultivation of *P. pastoris* X-33 harboring pPICZαC
131 with the mutated genes. Expression was conducted over three days in accordance with the
132 EasySelect™ Pichia Expression Kit (Invitrogen) protocol. Induction was carried out in 200 or 1000
133 mL BMMY (0.5 % methanol) while shaking at 28 °C. Protein synthesis was induced every 24 hours
134 by addition of methanol to a final concentration of 0.5 %. For high-density fermentation, *P.*
135 *pastoris* X-33 harboring pPICZαC with the mutated genes was fermented in a 5 L Sartorius Biostat
136 Aplus fermentor as described previously[9].
137 The His_6-tagged enzymes were purified from small-, or large-scale expression as described earlier[15]
138 using immobilized metal ion chromatography by loading the cell-free, concentrated enzyme-
139 containing cultivation medium to a 1 mL or 5 mL HisTrap™ HP column (GE Healthcare, Uppsala,
140 Sweden).

141

142 **2.4 Enzyme kinetic modeling**

143 A simple model describing the trans-sialylation of lactose using CGMP as a donor is presented in
144 Fig. 1. At the center of the model is the CI which can be formed from either of two michaelis
145 complex (MC). The MC with CGMP reacts to form CI (rate constant k1), whereas the MC with 3'SL
146 reacts to form CI (rate constant k-2). The CI can react with lactose to form the 3'SL MC (k2) or with
147 de-sialyllated donor to form the CGMP MC (k-1). Finally CI can react with water to release free SA
148 (k3). Infinitesimal amounts CI is formed from free SA and k-3 can therefore be neglected.
149 Product formation of both 3'SL and SA is dependent on the concentration of CI in the following
150 equations (under assumption of immediate product release):

$$V_{obs}^{3'SL} = \frac{d[3'SL]}{dt} = [CI] * k_2 - [3'SL] * k_{-2},$$

$$V_{obs}^{SA} = \frac{d[SA]}{dt} = [CI] * k_3 - [SA] * k_{-3}$$

When the reaction is observed at initial conditions, combination of the two equations enables a direct comparison of k2 and k3. Since

$k_{-3} \approx 0$ (due to a to large energy barrier) and [3'SL] = 0 at t = 0, it follows that:

$$\frac{V_{obs\_initial}^{SA}}{V_{obs\_initial}^{SA} + V_{obs\_initial}^{3'SL}} = \frac{[CI] * k_3}{[CI] * k_3 + [CI] * k_2} = \frac{k_3}{k_3 + k_2}$$

where $V_{obs\_initial}^{SA}$ and $V_{obs\_initial}^{3'SL}$ are the initial product formation rates. In the following, this expression is termed the initial hydrolytic specificity and will be used as a measure of how the trans-sialidase activity has changed. The product formation rates were estimated by taking the slope of the first three time points in the time course experiments shown in Fig. 4. This allowed calculation of the ratio of the rate constants k_2 and k_3 outlined above, and is used as a basis for discussion of mechanistic effects of the individual groups of mutations.

2.5 Trans-sialidase activity analysis

The donor substrate was purified as described previously[16] with the modification that all reaction mixtures were prepared in a 20 mM phosphate citrate buffer at pH 6.4. Lactose was added to the CGMP solution to final lactose and cGMP concentrations of 7.5 g/L and 50.6 g/L respectively. Enzyme preparations were diluted to equal concentrations of total protein and the concentration was confirmed by recording UV spectra of the diluted enzyme preparations (S1 Fig.). Enzyme

168 reactions were started by addition of 185 µl enzyme to 515 µl substrate mix (both preheated at 30
169 °C) and the resulting reaction mixture (15 µg/ml enzyme, 16 mM lactose, 37 g/L CGMP
170 corresponding to approximately 4 mM 3'-bound and 4 mM 6'-bound SA[17]) was incubated at 30 °C
171 with shaking (700 rpm). Reactions were stopped by heat inactivation at 90°C for 10 min before
172 they were transferred to a 5 kDa vivaspin filter (Sartorius) and centrifuged for 10 min at 5000 g
173 and 4 °C. Due to higher enzyme activity for some enzymes ($TrSA_{5mut}$, $TrSA_{6mut}$, $TrSA_{8mut}$, $TrSA_{12mut}$,
174 and $TrSA_{15mut}$), reactions were done using a lower enzyme concentrations (1.5µg/ml) with the
175 remaining variables held constant for these enzymes. Reactions were also performed using the
176 same enzyme preparations after heat-inactivation at 90 °C for 10 min. All reactions were done in
177 duplicates. The product concentrations in the reactions were measured by HPAEC-PAD.

178 **2.6 High-performance anion exchange chromatography (HPAEC-PAD)**

179 Separation and quantification of reaction products (3'SL and SA) from biocatalytic reactions were
180 carried out using a Dionex BioLC system consisting of GS50 gradient pumps, an ED50
181 electrochemical detector and an AS50 chromatography compartment coupled to an AS50 auto-
182 sampler (Dionex Corp., Sunnyvale, CA) as described previously[16].

183

184 # 3 Results and Discussion

185 **3.1 Design of TrSA mutants**

186 By integrating the findings of previous experimental and computational studies[1,14,2] we aimed to
187 design mutants of TrSA with high trans-sialidase activity. As a starting point, the mutant $TrSA_{5mut}$
188 was chosen as it was previously shown to contain the minimal amount of residues necessary for
189 conferring detectable trans-sialidase activity to the enzyme[14]. To do so, we grouped the mutations
190 based on expected functionality thereby obtaining three groups of mutations that were combined

191 in the $TrSA_{5mut}$ enzyme. The first group was termed "energetic mutants", and consisted of T38A,
192 F58N, and D284G, that based on the energy decomposition analysis were predicted to lower the
193 stability of the CI thus promoting trans-sialidase activity[2]. The second group was termed
194 "structural mutants" and consisted of I36L and G341A, that were predicted to improve the
195 mobility of the nucleophile[2]. However, due to the fact that all enzyme variants with the G341A
196 mutation was expressed at insufficient levels, the group was subsequently reduced to only include
197 I36L. The third group was termed "loop mutant" and consisted of the seven amino acid motif
198 (VTNKKKQ) that when introduced in $TrSA_{6mut}$ ($TrSA_{5mut}$/I36L) led to a reduced hydrolase activity
199 without negatively affecting the trans-sialidase activity [1]. Introduction of the different combination
200 of mutation groups into $TrSA_{5mut}$ led to the design of seven mutants as summarized in Fig. 2.
201
202 **3.2 Enzyme production in *P. pastoris***
203 To test the designed enzymes, we constructed genes encoding the mutants and expressed them in
204 the yeast *P. pastoris*. Expression levels of successfully expressed mutant variants were in the range
205 of 100 - 850 µg/l (determined by UV absorption after successful purification with no impurities
206 observed by SDS_PAGE) which was sufficient for the analyses performed in this study, and thus no
207 attempts were made at optimizing expression. As mentioned in the previous section, extremely
208 low or no expression was observed for variants carrying the mutation G341A alone or in
209 combination with I36L, respectively. In contrast, variants with only the I36L mutation expressed
210 well. A previous attempt to produce a mutant with both mutations in *E. coli* was also
211 unsuccessful[14]. Investigation of the TrSA[18] and the TcTS[19] crystal structures of the shows a
212 significant structural difference between the two enzymes. A serine residue (S340), adjacent to
213 the troublesome G341A mutation site, assume two different conformations in the TcTS and TrSA,

214 with phi/psi angles of -157.84/82.76 and -111.35/-126.14, respectively (see Fig. 3). Whereas the
215 serine in the wild-type TrSA is in an "allowed" conformation this is not the case in the TcTS. The
216 unstable conformation of the serine residue in TcTS is in a position to potentially contribute to the
217 mobility of the nucleophile Y342. Based on the structure, it appears possible that S340 will adopt
218 the same conformation in TrSA when the G341A mutation is introduced. Whereas the
219 destabilizing effect of G341A renders the TrSA mutants unstable it could be speculated that the
220 gradual evolution of the TcTS has allowed the serine to assume the "disallowed" conformation
221 through a series of stabilizing mutations.
222
223 **3.3 Trans-sialidase activity assessment**
224 To evaluate the produced TrSA mutant enzymes, we used the commercially interesting donor
225 substrate CGMP as donor and lactose as acceptor. We previously demonstrated that using high
226 acceptor concentrations leads to high trans-sialidase activity[9] even for enzymes with primarily
227 hydrolase activity. In order to highlight differences between the enzymes, lactose was used at a
228 lower concentration (donor:acceptor ratio of 1:4). With these conditions, time course experiments
229 were performed for all mutant enzymes and the effects of the introduced mutations on initial
230 donor consumption rate and initial hydrolytic specificity were analyzed. We initially optimized the
231 enzyme concentration using $TrSA_{6mut}$, but when applied to the other mutant enzymes, in case of
232 $TrSA_{5mut}$, $TrSA_{8mut}$, $TrSA_{12mut}$, and $TrSA_{15mut}$, we found that the hydrolytic activity was too high to
233 observe the initial 3'SL formation rate. Therefore the enzyme concentration was reduced 10-fold
234 for these variants. Since the observed rate constants could be susceptible to changes in enzyme
235 concentration, $TrSA_{6mut}$ was used as a center of reference between the two enzyme concentration
236 regimes, thus enabling comparison of all mutants and evaluation of effects of the individual

237 groups of mutation. $TrSA_{6mut}$ was chosen as the center of comparison between the fast enzymes
238 ($TrSA_{5mut}$, $TrSA_{8mut}$, $TrSA_{12mut}$, and $TrSA_{15mut}$) and the slow enzymes ($TrSA6_{mut}$, $TrSA8_{mut}$, $TrSA13_{mut}$,
239 and $TrSA16_{mut}$) as it showed the highest hydrolytic activity among the mutants containing the rate-
240 reducing active site mutation.
241
242 **3.4 Evaluation of time course experiments**
243 As displayed in Fig. 4 all of the mutants were analyzed in a time course experiment and all of the
244 mutants produce, in some cases slightly, more 3'SL than the benchmark enzyme $TrSA_{5mut}$ (Fig. 4A).
245 Since the study, by Pierdominici-Sotille et al., was conducted solely *in silico*, the effect of the
246 mutations has never been validated experimentally. It was suggested that $TrSA_{10mut}$ should have
247 trans-sialylation activity comparable to that of the TcTS. Expression of $TrSA_{10mut}$ was not possible.
248 The closest related expressed enzyme $TrSA_{9mut}$ (lacking the suggested G341A mutation which led
249 to insufficient expression) showed drastically increased trans-sialidase activity but also a high
250 degree of hydrolytic activity and thus product degradation was observed. $TrSA_{5mut}$, $TrSA_{6mut}$ and
251 $TrSA_{13mut}$ have all been characterized previously, but this is the first study in which these mutants
252 have been compared using same experimental conditions. $TrSA_{6mut}$ was first described as a mutant
253 with improved trans-sialidase activity compared to the $TrSA_{5mut}$. In that study, where μ-molar
254 product concentrations were analyzed using 3'SL as donor and C^{14}-labeled lactose as acceptor,
255 $TrSA_{6mut}$ exhibited 11.3% the trans-sialidase activity of the TcTS compared to 0.9% for $TrSA_{5mut}$[14].
256 Whereas this finding cannot be directly confirmed or disproved by our experiments (partly due to
257 the differences between the two experimental setups), we observe that where $TrSA5_{mut}$ shows
258 higher 3'SL formation (relative to enzyme concentration). Although this finding seem contradictory
259 to the findings of Paris et. al. it might be possible to explain since the trans-sialidase activity

260 reported in this early study is based on a single time point. Thus product hydrolysis could have had
261 great influence on the reported trans-activity. What is clear from the direct comparison of $TrSA_{5mut}$
262 and $TrSA_{6mut}$ is that the product to hydrolysis ratio is greatly enhanced for $TrSA_{6mut}$ compared to
263 $TrSA_{5mut}$. As mentioned we have previously used the $TrSA_{6mut}$ for production of sialylated glycans
264 and we have reported on the superior mutant $TrSA_{13mut}$. In our previous work, $TrSA_{6mut}$ and
265 $TrSA_{13mut}$ were compared based on a fluorometric time course assay using an artificial acceptor
266 substrate[1]. In this study, using a more relevant acceptor substrate we confirmed that the
267 hydrolytic activity of $TrSA_{13mut}$ is reduced whereas the trans-sialidase activity remained unaffected.
268 However, comparing the reaction model (Fig. 1) and the overview of energy barriers showed in
269 Fig. 5 it seems convincing that this is a result of k1 decreasing while k2 increase accordingly as
270 results of the changed energy requirements at the transition states.
271 With improved 3'SL production in mind, it is hard to directly compare $TrSA_{6mut}$, $TrSA_{9mut}$ and
272 $TrSA_{13mut}$, based on time-course experiment (Fig. 4B, D, and F), but there is no doubt that we by
273 combining the mutations in $TrSA_{16mut}$ (Fig. 4H) have created a superior mutant displaying
274 dramatically better product to hydrolysis ratio. Furthermore, time course experiments are
275 presented for the mutants $TrSA_{8mut}$ and $TrSA_{15mut}$ carrying the energetic mutations without the
276 structural mutation (Fig. 4C and G). Whereas $TrSA_{8mut}$ does not perform as well as either of the
277 discussed mutants $TrSA_{15mut}$ seem to compete with $TrSA_{16mut}$. In the TrSA15mut time course the
278 concentration of 3'SL reaches a plateau at a lower concentration than the highest 3'SL
279 concentration measured for $TrSA_{16mut}$. Keeping in mind, however, that the concentration of
280 $TrSA_{16mut}$ used for the time study is 10 times that of $TrSA_{15mut}$, it could be argued that Tr_{15mut} is the
281 superior enzyme. $TrSA_{12mut}$ with only the loop mutation (Fig. 4E) was an improvement over
282 $TrSA_{5mut}$ but showed inferior performance compared to the other mutants.

283

284 **3.5 Effect of mutation groups**

285 To discuss the effect of the individual mutation groups initial hydrolytic specificity and initial donor
286 consumption rate (Table 2) will be held up against the reaction model (Fig. 1) reaction energy
287 profile presented in Fig. 5. From the model it is clear that the efficiency of the trans-sialylation
288 reaction heavily depend on what attacks the CI. The rate constant determining hydrolysis is given
289 by k3 and it is clear that the consequence of complete elimination of hydrolysis will be a reaction
290 which will result in the natural equilibrium determined by k1, k-1, k2 and k-2. Since CI formation
291 from donor and product are similar reactions standard enzyme kinetics states that k1 and k-2 are
292 of similar sizes. The same is the case for k2 and k-1, but whereas it seems crucial to focus on k3 for
293 optimization of trans-sialidase activity, it turns out that optimizing k2 is just as crucial. When trans-
294 sialylation is at all possible it must mean that water is very restricted from the active site as its
295 molar concentration is approximately 3400 times that of lactose. Thus the observed hydrolysis by
296 $TrSA_{5mut}$ is believed to be correlated to the hampered trans-sialidase activity (a small k2) and
297 increasing k2 will therefore effectively lead to decreased hydrolysis (a smaller k3).

298

299 *Energetic*

300 The energetic mutation when introduced on its own reduces the initial donor consumption by 3.2
301 fold and the initial hydrolytic specificity by 2.6 fold. This can be explained by an alteration of the
302 energy profile for CI formation (Fig. 5). When the profile shifts towards the TcTS profile, k1 is
303 reduced due to an increased energy barrier for CI formation whereas k2 is increased as the energy
304 barrier for this step is decreased as a result of the CI destabilization. Whereas the decreased initial
305 donor consumption speed is an overall result of a decreased k1, the reduced initial hydrolytic

306 specificity is a result of an increased k2, which in turn leads to a decrease in k3 as hydrolysis is
307 favored by reduced trans-sialylation[2].

308 *Structural*

309 The structural mutation has the greatest effect on both initial donor consumption rate and initial
310 hydrolytic specificity. Compared to the benchmark enzyme $TrSA_{5mut}$, with no additional
311 background the initial donor consumption rate is reduced 16.1 fold. Such a drastic decrease (or a
312 decrease at all) is surprising because the mutation is suggested to increase the flexibility of the
313 nucleophile – a function that does not immediately relate to reduced activity. If the rearranged
314 nucleophile has had an effect on the energy profile, destabilizing the CI while increasing the
315 barrier at the transition state between MC and CI thereby decreasing k1, a decrease in the initial
316 donor consumption rate could be the result, but it does not seem convincing that this effect
317 should be ~ 6 fold larger than the effect of the energetic mutations. An alternative explanation is
318 that by rearranging the nucleophile, the CI has become less susceptible to an attack from water
319 and or an acceptor which would explain the decrease of the initial donor consumption rate as a
320 decrease of k3 and/or k2 respectively. Looking at the initial hydrolytic specificity it becomes clear
321 that both 3'SL and SA formation is drastically decreased. So is, however, the hydrolytic specificity
322 of the enzyme, by 3.1 fold. This decrease is believed to be due to a drastic decrease of k3,
323 supporting the notion that the structural mutation has had the effect of sterically protecting the CI
324 against hydrolysis. Moreover, if the hypothesis of a sterically protected CI is true it cannot be
325 excluded that CI formation is also hampered due to a less accessible active site, resulting in a
326 smaller k1 value.

327 *Loop mutations*

328 Looking at the loop mutations by themselves it seems convincing that they have had an effect on
329 k2. While the initial hydrolytic specificity is reduced by 1.4 fold, the initial donor consumption is
330 relatively unchanged. The loop mutation has previously been suggested to influence the water
331 network surrounding the active site, thereby hindering water from attacking the CI. Such
332 hindrance would result in a smaller k3 which in turn could lead to a decreased hydrolytic
333 specificity. A smaller k3 would however lead to a simultaneous decrease in initial donor
334 consumption. Accordingly since no such effects is observed the effect of the mutation must in part
335 be ascribed to an increased k2. Looking at Fig. 4 (A and E) a simultaneous increase and decrease of
336 k2 and k3 seem convincing. The increased k2 value is most easily explained by a decreased energy
337 barrier between the CI and the MC' at the transition state as a result of a destabilization of the CI,
338 which would also contribute to a decreased k1, whereas the decrease in k3 can be ascribed to the
339 increase in k2. When no decrease in initial donor consumption rate is observed as result of a
340 smaller k1 it is because the increased k2 value eventually enables the enzyme to be recycled faster
341 (which in turn leads to a higher MC concentration).

342 It is speculated that since the loop structure in itself is not in contact with the substrate it may
343 influence the surroundings or residues influencing the CI stability, among which some may be
344 included in the QM subsystem defined by Pierdominici-Sottile et al.[2] The closest of the QM
345 subsystem residues is D96, which is also the residue which leaves better room for improvement as
346 it is the only residue of the QM subsystem which remains to have a net stabilizing effect in the
347 $TrSA_{10mut}$.

348 *Supporting the notions by combinatory mutations*

Further support for the hypotheses discussed above was sought by investigating the mutations in combination with each other. For this part of the discussion we will consider a combination of two mutations as an introduction of one of the mutations into the background of the other mutation. For clarity it should be noted that this introduction is merely intellectual as naturally only one combinatory mutant has been expressed. The introduction of a mutant into a mutational background is simply used to understand the effect of the mutation with that particular background as the reference point.

The hypothesis that the energetic mutations shift the energy profile towards the TcTS variant is further supported by evaluation of the energetic mutation in combination with the remaining mutations. When the energetic mutations are introduced into the structural mutational background it is seen that the initial hydrolytic specificity is reduced while the initial donor consumption remain unchanged. While a coincidence leads to identical fold changes in initial donor consumption and initial hydrolytic specificity as for the individually introduced loop mutation, the argument that follows is the same and suggests an increased k2 value together with a decreased k1 value due to a destabilized CI as seen in the TcTS-like energy profile.

Furthermore there is a strong synergistic effect between the energetic mutation and the loop mutations which support the hypothesized mechanisms of both mutations. The energetic mutations and the loop mutations reduce the initial hydrolytic specificity is reduced by 2.6 fold and 1.4 fold respectively, when introduced individually into the benchmark enzyme ($TrSA_{5mut}$ vs. $TrSA_{8mut}$ and $TrSA_{5mut}$ vs. $TrSA_{12mut}$). However, when the energetic mutation is introduced to the loop mutational background and *vice versa* reductions of 8.0 fold and 4.2 are observed respectively. This strongly supports the hypothesis that the loop mutations affect the surroundings

371 of the CI stabilizing residues in the enzymes. Besides a hypothesized effect on D96 (or another CI
372 stabilizing residue) the loop mutation is believed to strongly affect one or more of the residues
373 introduced energetic mutations creating the synergistic effect of the two mutations. However an
374 energy decomposition analysis, preferably based on the crystal structure of the mutants, would be
375 necessary to establish exactly how the loop interacts with theses residues.

376

377 The hypothesized mechanism of the structural mutation, that it results in a sterically protected CI,
378 is further supported when the structural mutation is introduced into the energetic mutations
379 background. Here the effect on initial donor consumption is less pronounced with a reduction of
380 4.9 fold. This suggests that while the CI becomes more protected from water by introduction of
381 the structural mutation, the lower energy barrier for MC' formation in the energetic mutation
382 background (leading to an increased k2), has reduced the effect of the structural mutation on
383 initial donor consumption. In the energetic mutational background, the structural mutation it is
384 also observed to have a smaller effect (although still diminishing) on the initial hydrolytic
385 specificity. This supports the above argument as steric protection against hydrolysis has less of an
386 impact on k3 when the CI-MC' energy barrier is reduced.

387 In contrast there is a large synergistic effect, with regard to initial donor consumption rate, of
388 introducing the structural mutation in the loop mutational background. The synergistic effect is
389 not surprising as the loop mutation was suggested to have had a diminishing effect on the k1 value
390 (as discussed above) by increasing the energy barrier at the MC-CI transition state, and the
391 combined effect of the two mutations – e.g. simultaneous steric hindrance and an increased MC-
392 CI energy barrier – would not be limited to the sum of the two. On the other hand a similar fold

393 reduction in initial hydrolytic specificity can be observed which make sense if a steric protection of
394 the CI is taking place. The steric protection will change the k3 value as a result of a reduced
395 likelihood of a water molecule finding its way the CI and in that regard the effect of the mutations
396 is additive to an effect caused by a changed CI formation energy profile. Furthermore, when
397 introducing the structural mutation to the mutational backgrounds of the energetic mutations
398 with or without the loop mutations (TrSA15mut vs. TrSA16mut and TrSA8mut vs. TrSA9mut) the
399 patterns are preserved. Thus the additive effect on initial hydrolytic specificity sees fold reductions
400 of 1.4 fold and 1.7 fold respectively, whereas the non-additive (synergetic) effect on initial donor
401 consumption grants fold reductions of 19.4 fold and 4.9 fold respectively.

402 **3.6 The best mutants**

403 Based on the initial hydrolytic specificity, two enzymes stand out as drastically improved,
404 exhibiting initial hydrolytic specificity at less than 10% but whereas the enzyme with the lowest
405 initial hydrolytic rate, $TrSA_{16mut}$, has the initial donor consumption rate-inhibiting active site
406 mutation incorporated, this is not the case for the other mutant, $TrSA_{15mut}$. Thus, depending on
407 the goal of a specific trans-sialylation reaction, each of the two enzymes is very potent. When the
408 goal is substrate utilization $TrSA_{16mut}$ will outperform $TrSA_{15mut}$, whereas $TrSA_{15mut}$ is a much faster
409 enzyme. The overall difference between the two enzymes is believed to be an active site more
410 sterically shielded in TrSA16mut, most likely resulting in a decrease of all rate constants used to
411 describe the reaction.

412

413 # 4. Conclusions

414 We have introduced mutations suggested by three different studies into a benchmark enzyme, the
415 engineered $TrSA_{5mut}$ and evaluated the effect of the mutations. A total of 12 mutations were
416 suggested, which were split into 3 groups, due to the suggested mechanistic properties suggested
417 in three studies conducted by Paris et al., Jers et al., and Pierdominici-Sotile et al. Two of the
418 mutations were as previously reported not possible to introduce together with successful
419 expression and we found that one of the two mutations consistently led to insufficient expression.
420 In a time study, resembling a genuine HMO production process, using the commercially viable
421 substrate CGMP, seven mutants were investigated to evaluate the effects of the individual
422 mutation groups and to evaluate whether an engineered sialidase with improved trans-sialidase
423 activity could be produced. All of the expressed mutations contributed to enhanced trans-sialidase
424 activity in that they enhanced the product specificity of the enzymes. A single point mutation
425 responsible for improving trans-sialidase activity, through increasing active site nucleophile
426 mobility, reduced the activity of enzymes dramatically but two mutants (one with and one without
427 the activity reducing mutation) proved to be superior trans-sialidases compared to engineered
428 sialidases, previously experimentally verified. As a consequence not one but two of the mutants,
429 depending on the aim of use, are each in its own right the best engineered trans-sialidase to date.
430

431 ## Acknowledgements

432 Funding of work

References

1. Jers, C. *et al.* Rational design of a new Trypanosoma rangeli trans-sialidase for efficient sialylation of glycans. *PLoS One* **9,** e83902 (2014).
2. Pierdominici-Sottile, G., Palma, J. & Roitberg, A. E. Free-energy computations identify the mutations required to confer trans-sialidase activity into Trypanosoma rangeli sialidase. *Proteins* **82,** 424–35 (2014).
3. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22,** 1147–62 (2012).
4. Angeloni, S. *et al.* Glycoprofiling with micro-arrays of glycoconjugates and lectins. *Glycobiology* **15,** 31–41 (2005).
5. Sela, D. A. & Mills, D. A. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol.* **18,** 298–307 (2010).
6. Lin, A. E., Autran, C. A., Espanola, S. D., Bode, L. & Nizet, V. Human milk oligosaccharides protect bladder epithelial cells against uropathogenic Escherichia coli invasion and cytotoxicity. *J. Infect. Dis.* **209,** 389–98 (2014).
7. Castillo-Courtade, L. *et al.* Attenuation of food allergy symptoms following treatment with human milk oligosaccharides in a mouse model. *Allergy* **70,** 1091–102 (2015).
8. Braegger, C. *et al.* Supplementation of Infant Formula With Probiotics and/or Prebiotics: A Systematic Review and Comment by the ESPGHAN Committee on Nutrition. *J. Pediatr. Gastroenterol. Nutr.* **52,** 238–250 (2011).
9. Michalak, M. *et al.* Biocatalytic production of 3′-sialyllactose by use of a modified sialidase with superior trans-sialidase activity. *Process Biochem.* **49,** 265–270 (2014).
10. Holck, J. *et al.* Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by Trypanosoma cruzi trans-sialidase. *N. Biotechnol.* **31,** 156–65 (2014).
11. Zeuner, B., Jers, C., Mikkelsen, J. D. & Meyer, A. S. Methods for improving enzymatic trans-glycosylation for synthesis of human milk oligosaccharide biomimetics. *J. Agric. Food Chem.* **62,** 9615–31 (2014).
12. Pereira, M., Zhang, K. & Gong, Y. Invasive phenotype of Trypanosoma cruzi restricted to a population expressing trans-sialidase. *Infect. …* **64,** 3884–3892 (1996).
13. Pontes-de-Carvalho, L. C., Tomlinson, S. & Nussenzweig, V. Trypanosoma rangeli sialidase lacks trans-sialidase activity. *Mol. Biochem. Parasitol.* **62,** 19–25 (1993).
14. Paris, G. *et al.* A sialidase mutant displaying trans-sialidase activity. *J. Mol. Biol.* **345,** 923–34 (2005).
15. Zeuner, B. *et al.* Optimizing the biocatalytic productivity of an engineered sialidase from Trypanosoma rangeli for 3′-sialyllactose production. *Enzyme Microb. Technol.* **55,** 85–93 (2014).
16. Luo, J. *et al.* An integrated membrane system for the biocatalytic production of 3′-sialyllactose from dairy by-products. *Bioresour. Technol.* **166C,** 9–16 (2014).
17. Saito, T. & Itoh, T. Variations and distributions of O-glycosidically linked sugar chains in bovine kappa-casein. *J. Dairy Sci.* **75,** 1768–74 (1992).
18. Amaya, M. ., Buschiazzo, A., Nguyen, T. & Alzari, P. M. The high resolution structures of free and inhibitor-bound Trypanosoma rangeli sialidase and its comparison with T. cruzi trans-sialidase. *J.Mol.Biol.* **325,** 773–784 (2003).

19. Amaya, M. F. *et al.* Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase. *Structure* **12,** 775–84 (2004).

Tables

Table 1: List of primers

Name	Sequence
T38A(I36L)_fwd	GGTTCATTCATTTAGATTACCAGCTATCGTTAACGTAGATGGAGT
T38A(I36L)_rev	ACTCCATCTACGTTAACGATAGCTGGTAATCTAAATGAATGAACC
F58N_fwd	CTGATGCCAGATATGAGACATCAAACGACAACTCCTTTATCGAA
F58N_rev	TTCGATAAAGGAGTTGTCGTTTGATGTCTCATATCTGGCATCAG
D284G_fwd	CAACTTCCAATCAACCCGGTTGTCAGAGTTCATTCGT
D284G_rev	ACGAATGAACTCTGACAACCGGGTTGATTGGAAGTTG
G341A_fwd	ATTGGTGATGAAAACAGTGCTTACTCTTCCGTCCTATAC
G341A_rev	GTATAGGACGGAAGAGTAAGCACTGTTTTCATCACCAAT
L36I_fwd	CGTGTGGTTCATTCATTTAGAATACCAACTATCGTTAACGTAG
L36I_rev	CTACGTTAACGATAGTTGGTATTCTAAATGAATGAACCACACG
T38A(I36)_fwd	CGTGTGGTTCATTCATTTAGAATACCAGCTATCGTTAACGTAGATGG
T38A(I36)_rev	CCATCTACGTTAACGATAGCTGGTATTCTAAATGAATGAACCACACG

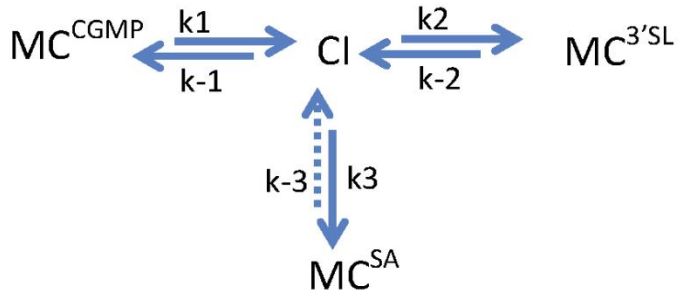
Table 2: Initial donor consumption rate and hydrolytic specificity

The effects of the individual mutations are given as a fold reduction in initial donor consumption and initial hydrolytic specificity (see Materials and methods section). The effect of each group of mutation is in the

table considered against each available mutational background. Thus the donor consumption rate is reduced 39.6 fold when the structural mutation was introduced in the mutant containing the loop mutation ($TrSA_{12mut}$).

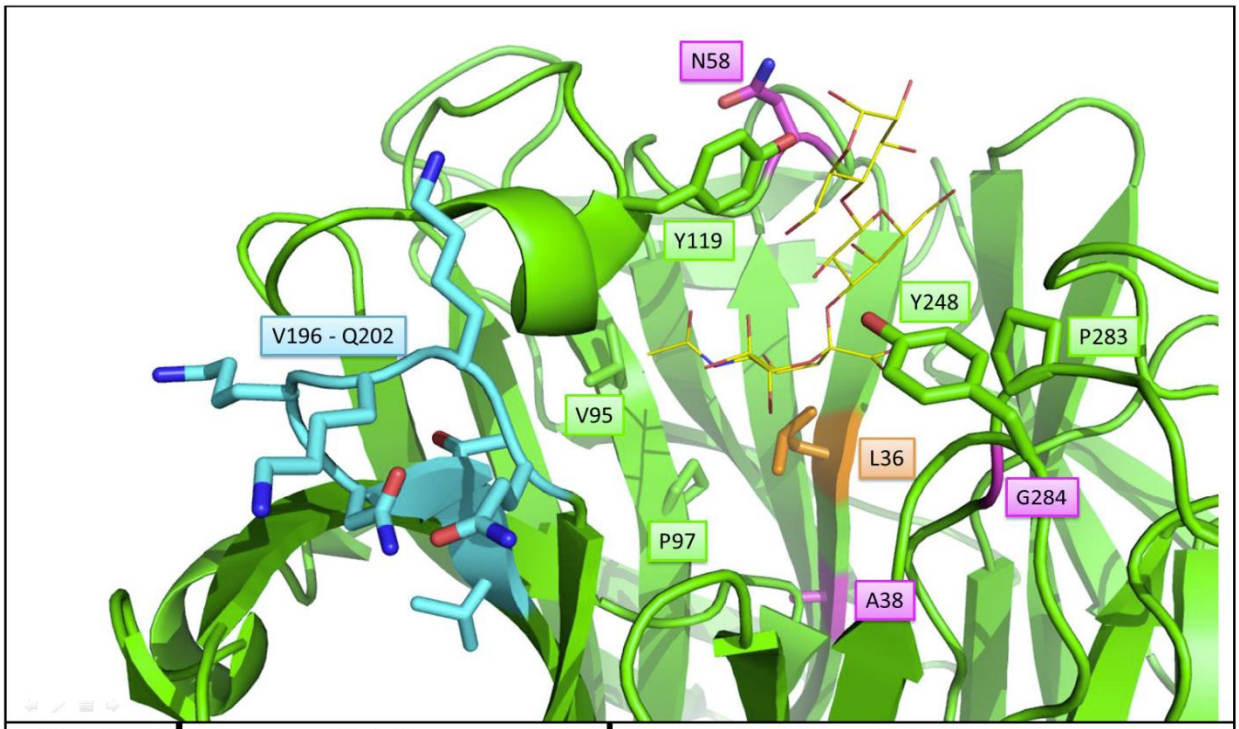
Introduced Mutation	Mutational background	Comparison	Initial donor consumption rate			Initial hydrolytic specificity		
			-mut	+ mut	Fold reduction	-Mut	+Mut	Fold reduction
Energy	None	Tr5 vs Tr8	100.0%	31.6%	3.2	100.0%	38.0%	2.6
	Structural	Tr6 vs Tr9	6.2%	6.4%	1.0	32.4%	22.8%	1.4
	Loop	Tr12 vs Tr15	98.6%	74.6%	1.3	71.9%	9.0%	8.0
	Struc. + loop	Tr13 vs Tr16	2.5%	3.9%	0.6	21.2%	6.4%	3.3
Structural	None	Tr5 vs Tr6	100.0%	6.2%	16.1	100.0%	32.4%	3.1
	Loop	Tr12 vs Tr13	98.6%	2.5%	39.6	71.9%	21.2%	3.4
	Energetic	Tr8 vs Tr9	31.6%	6.4%	4.9	38.0%	22.8%	1.7
	Loop + Ener.	Tr15 vs Tr16	74.6%	3.9%	19.4	9.0%	6.4%	1.4
Loop	None	Tr5 vs Tr12	**100.0%**	**98.6%**	1.0	**100.0%**	**71.9%**	1.4
	Structural	Tr6 vs Tr13	**6.2%**	**2.5%**	2.5	**32.4%**	**21.2%**	1.5
	Energetic	Tr8 vs Tr15	**31.6%**	**74.6%**	0.4	**38.0%**	**9.0%**	4.2
	Struc. + Ener.	Tr9 vs Tr16	**6.5%**	**3.9%**	1.7	**22.8%**	**6.4%**	3.6

509 # Figures



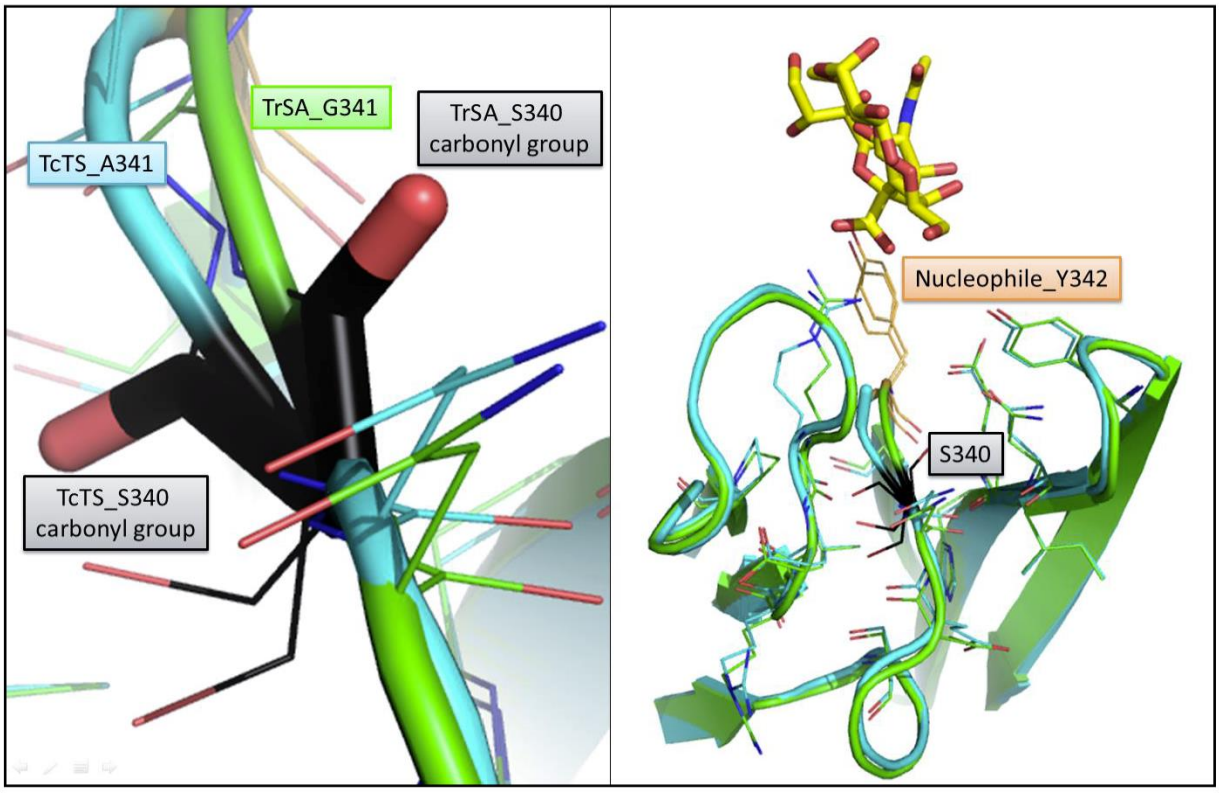
510

511 **Figure 1: Model for formation and deformation of CI**



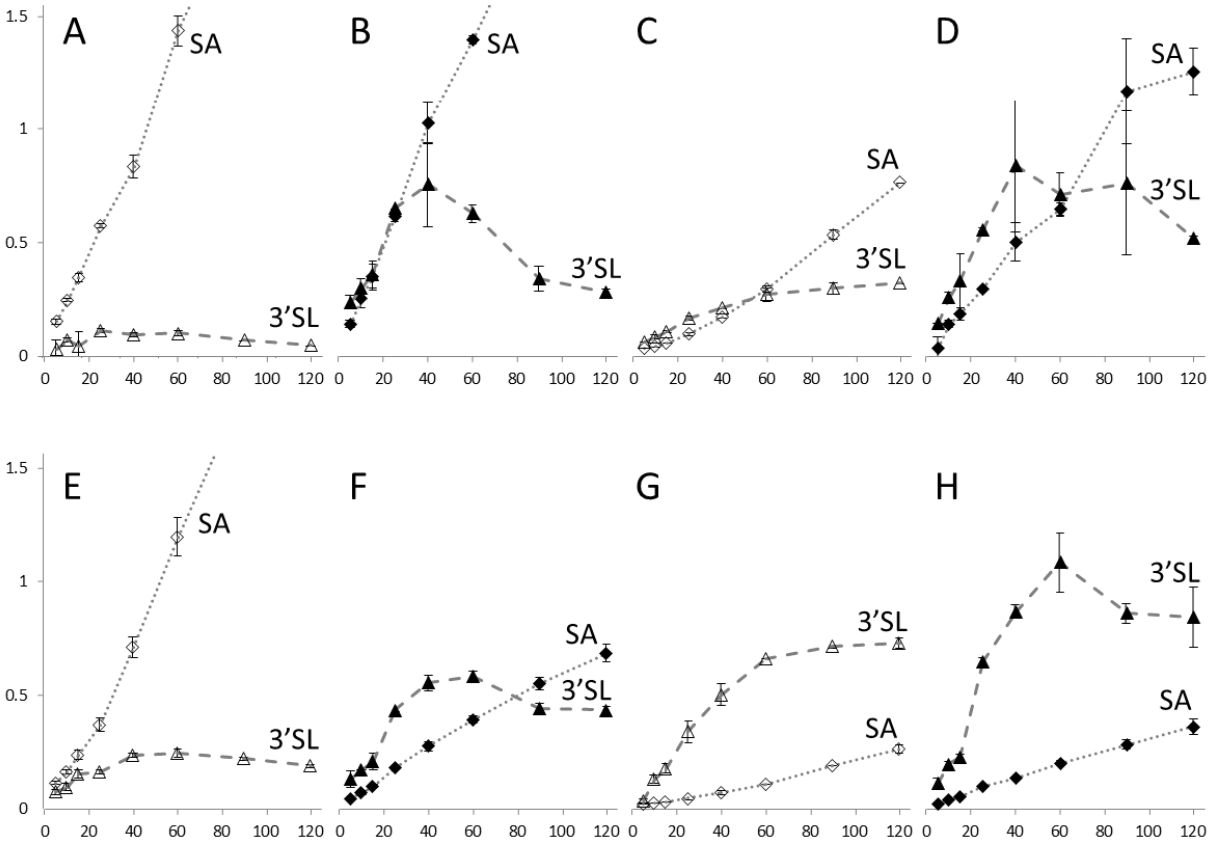
Mutation groups	Mutations within groups	Mutants							
		$TrSA5_{mut}$	$TrSA6_{mut}$	$TrSA8_{mut}$	$TrSA9_{mut}$	$TrSA12_{mut}$	$TrSA13_{mut}$	$TrSA15_{mut}$	$TrSA16_{mut}$
$TrSA5_{mut}$	M95V, A97P, S119Y, G248Y, Q283P	X	X	X	X	X	X	X	X
Energetic	T39A, F59N, D285G			X	X			X	X
Structural	I36L		X		X		X		X
Loop	Loop "VTNKKKQ"					X	X	X	X

512 **Figure 2: Mutations introduced to the TrSA**



513

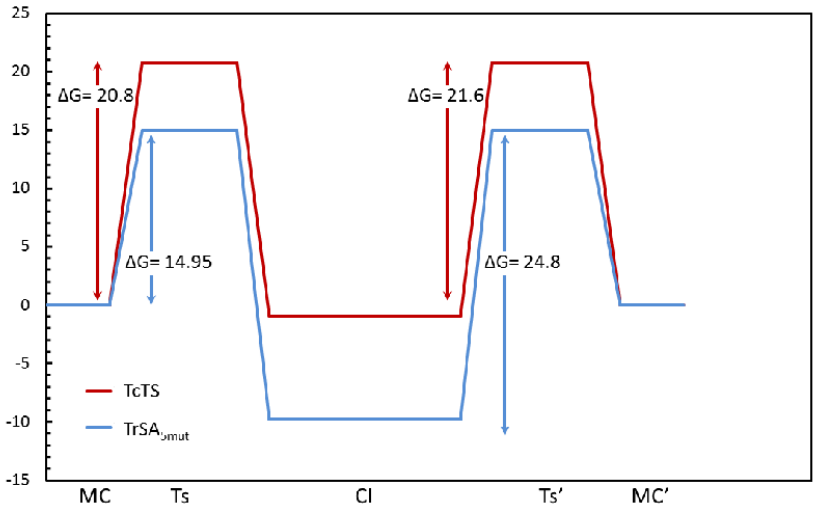
514 **Figure 3. Inspection of 3D structures of TrSA and TcTS around S340**



519

520 **Figure 4: Time course experiments for 3'SL production with novel mutants**

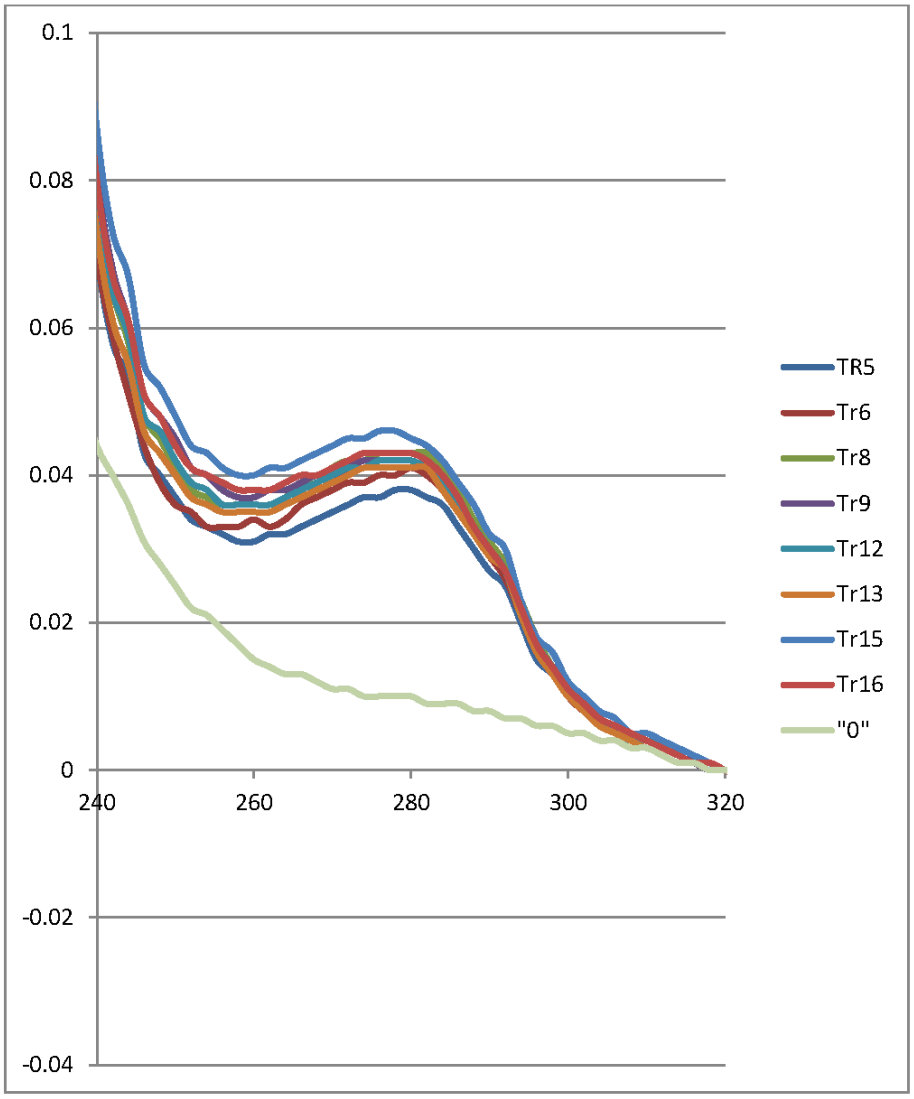
521



522

523 **Figure 5: Energy profile for CI formation and deformation**

524



Supplementary figure S1: Dilution of enzymes for reactions.

Figure captions

Figure 2: Model for formation and deformation of CI

The model describes the reaction based on five rate constants (considering the contribution for k-3 to be infinitesimal due to an unfavorable energetic barrier). At the center of the model is the CI which is formed through k1 and k-2 whereas it is consumed through k2, k3, and k-1. Due to donor and acceptor similarity it is expected that k1≈k-2 and k2≈k-2, whereas k3 is not directly related to any of the other constants. MC^{CGMP}, MC^{SA}, and MC^{SA} denotes the MC formed with CGMP, 3'SL and free sialic acid.

Figure 2: Mutations introduced to the TrSA

The eight enzymes investigated in this study were constructed from the TrSA by systematic combination of four groups of mutations. In addition to a group consisting of the five mutations reported to be essential to trans-sialylation activity (by Paris el al.), the mutations were grouped based on their suggested mechanistic contribution to trans-sialidase activity. The figure also grants a graphical overview of the 16 residues which in different combinations were mutated in TrSA during construction of different mutants and 3'SL is included to indicate the orientation of the active site. Thus the figure depicts the Micaelis complex of 3'SL with TrSA16mut. The residues are shown in their mutated form and are colored accordingly to the four groups: $TrSA_{5mut}$-mutations (green), energetic (magenta), structural (orange), and loop (cyan).

Figure 3. Inspection of 3D structures of TrSA and TcTS around S340

Mutation of G341 in the TrSA led to insufficient expression. The figure shows the conformation of the carboxylic oxygens of S340 adjacent to G341 in each of the two enzymes TcTS and TrSA (A). In TcTS the residue assume a conformation with a disallowed phi/psi angle set and it is believed that the mutation of the adjacent residue (G341A) leads to a similar phi/psi angle set of the S340 in the TrSA since no other differences can be observed in the structures around S340 (B). Whereas the disallowed conformation of S340 leads renders TrSA mutants that cannot be expressed it is speculated that years of evolution of has allowed the destabilization caused by S340 in the TcTS.

Figure 4: Time course experiments for 3'SL production with novel mutants

Depicted are the time course curves for the eight enzymes A) $TrSA_{5mut}$, B) $TrSA_{6mut}$, C) $TrSA_{8mut}$, D) $TrSA_{9mut}$, E) $TrSA_{12mut}$, F) $TrSA_{13mut}$, G) $TrSA_{15mut}$, and H) $TrSA_{16mut}$. In the top row are (A-D) mutants without the loop mutation, bottom row (E-H) with loop mutation, left hand side (A,B,E,F) without energetic mutations, right hand side (C,D,G,H) with energetic mutations, hollow markers (A,C,E,G) without structural mutation, and black markers (B,D,F,H) with structural mutations. The effect of the structural mutation on enzyme activity is more pronounced than it appears since these mutants were analyzed using 1.5 μg/ml enzyme compared to 15 μg/ml of the remaining enzymes. From the graphs it is clear that the two most potent enzymes are $TrSA_{15mut}$, $TrSA_{16mut}$, whereas none of the suggested mutations individually confer total trans-sialidase activity. Higher resolution plots of all of the time studies can be found in supplementary data as some of the mutants (A,B and E) produce SA at a high rate and the graphs exit the plots.

Figure 5: Energy profile for CI formation and deformation

The figure is modified from Pierdominici-Sottile et al[2]. The figure shows the energy barriers at transition states between MC, CI and MC with acceptor (MC') and it is seen that the energy barrier for CI formation is larger for TcTS than for TrSA5mut, whereas MC' formation barrier is smaller due to a less stable CI. The reduced stability and smaller energy barrier should result in quicker turnover and thus less exposure to water resulting in less hydrolysis.

Supplementary figure S1: Dilution of enzymes for reactions.

UV spectra of the diluted enzyme samples before added to reactions. On SDS-page no contaminations were observed (data not shown) and thus the spectra in the figure show that enzymes were successfully diluted to equal concentrations.

Chapter 5: Separation of 3'-sialyllactose and lactose by nanofiltration: A trade-off between charge repulsion and pore swelling induced by high pH

5.1 Motivation

The process for production of sialylated HMOs set up by the group relies of the use of engineered (trans-) sialidases (from hydrolytic wild type enzymes). The optimized reaction conditions rely on high concentrations of acceptor (Lactose) to avoid hydrolysis and product concentrations have never reached more than 4 mM compared to an acceptor concentration of approximately 350 mM. Furthermore the standard concentration of CGMP used as donor molecule for the reactions is 40 g/l (corresponding to 4 mM 3'-bound SA). Hence a strategic purification scheme is necessary to economically optimize the HMO production process with the following three main goals: 1) obtaining product of suitable purity applicable for infant formula supplementation, 2) separation of the de-sialylated CGMP which retain several of its properties as a food additive, and 3) separation of lactose to be recycled as acceptor substrate in the reaction. This study primarily sought to achieve goal 1 whereas goals 2 and 3 were included in the study described in Chapter 6. A strong motivating factor for this study was that large quantities of pure HMOs were needed in order to research the functionality of the produced HMOs, in relation to human consumption. HMO functionality is however outside the scope of this PhD project and the scientific work conducted is therefore framed in the context of production throughout the published material.

5.2 Hypotheses and objectives

The objective of this study was well defined and entailed purification of 3'SL at sufficient purity (99%) suitable not only for application in infant formula but for scientific evaluation of functional properties. Besides the high grade of purity it was important that any impurities would be immunological inert and non-toxic to microorganisms.

The main hypothesis relevant to the overall goal of the PhD project was concerning the objective of separating 3'SL from lactose achieving 99% purity of 3'SL. The hypothesis was constructed on the basis of the experimental considerations and reads:

Hypothesis 5.1: Since NF efficiency can be heavily affected by the membrane zeta potential (charge), and since the charge difference between lactose and 3'SL is large, NF it will be an efficient tool for separating 3'SL and lactose despite similar size of the molecules (from a high molecular weight NF perspective).

5.3 Experimental considerations

The initial approach to purification was column chromatography, on an anion exchange resin. This has previously been applied successfully for 3'SL purification in a lab-setting, but using the unfavorable (due to toxicity) eluent acetonitrile. To avoid using this eluent it was suggested that NaCl could be used instead, which turned out to be the case. The motivation for identification of a different way of purifying the

product was, however, mainly due to a forthcoming task of up-scaling the reaction to 50l scale. Here purification would be a considerable bottleneck as no large scale column was available. Combining this with a subsequent NF step being an inescapable consequence of elution by NaCl, the idea of using NF to separate 3'SL and lactose was considered. NF had in the first place been ruled out as the size difference between 3'SL and lactose was not evaluated to be large enough to ensure efficient separation. Due to the high concentration of lactose compared to 3'SL lack of efficient separation would thusly result in an inacceptable loss of product.

The idea to revisit the possibility of NF separation of 3'SL and lactose was sparked by failure to separate CGMP and enzyme from the remaining reaction components in the preceding ultrafiltration (UF) step. In large scale the first attempt of separation resulted in no permeation of 3'SL across the membrane despite a MWCO of 10000 Da. The cause of 3'SL retention was speculated to be CGMP fouling rendering negatively charged surface efficiently repelling the 3'SL. Thus it was decided to set up experiments using a charged NF membrane to efficiently separate 3'SL and lactose.

5.4 Conclusions

The application of NF for separation of 3'SL and lactose was successful using a suitable negatively charged membrane. Despite the membrane having, reportedly (by the manufacturer), a molecular weight cutoff of 600-800 Da – a MWCO which should allow for 3'SL permeation – high 3'SL retention was observed (~100%). It was evaluated that the retention was due to charge difference and mathematical modeling showed that high purity could be achieved (with an acceptable loss of product) by dia-filtration. To confirm the model 10 rounds of dia-filtration was carried out using a model solution (pure chemicals) and a "real" solution (the product mixture of a trans-sialylation reaction). Both dia-filtration experiments outperformed the model suggesting that NF based on solute charge differences is indeed a suitable method for separation of the two reaction products.

5.5 Paper IV

Separation and Purification Technology 138 (2014) 77–83

Contents lists available at ScienceDirect

Separation and Purification Technology

journal homepage: www.elsevier.com/locate/seppur

Separation of 3′-sialyllactose and lactose by nanofiltration: A trade-off between charge repulsion and pore swelling induced by high pH

Rune T. Nordvang, Jianquan Luo, Birgitte Zeuner, Rasmus Prior, Mads F. Andersen, Jørn D. Mikkelsen, Anne S. Meyer, Manuel Pinelo *

Technical University of Denmark, Department of Chemical and Biochemical Engineering, Center for BioProcess Engineering, Søltofts Plads Bygning 229, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form 9 October 2014
Accepted 10 October 2014
Available online 22 October 2014

Keywords:
Nanofiltration
Prebiotics
3′-Sialyllactose
Electrostatic repulsion
Pore swelling

ABSTRACT

Separation of 3′-sialyllactose (SL) and lactose is an essential final step for the production of the next generation of infant formulas containing sialyllated prebiotics. Due to the difference in molecular weight (MW) between SL and lactose and the charge density of SL, nanofiltration could provide a rapid, inexpensive alternative for the separation of SL and lactose compared to traditional chromatography. The performance of four commercial nanofiltration membranes (NF45, DSS-ETNA01PP, NTR-7540 and NP010) for the separation of SL and lactose was assessed at various pH. The difference in retention between SL and lactose was only significant in the NP010 and NTR-7450 membranes, whereas the NF45 and DSS ETNA01PP membranes exhibited either too high lactose retention (i.e. insufficient separation) or too low SL retention (i.e. losing the target SL compound), respectively. Operation at increased pH did not affect SL retention significantly. The expected increase in retention levels of SL at high pH – due to repulsion between the negative charge of the membrane and the charged SL – was apparently offset by pore swelling of the NF membranes at high pH. The water permeability was measured before and after a membrane was used for filtration of a mixture of lactose and SL. For the NP010 and DSS-ETNA membranes, the decline in water permeability was lower when the experiments were conducted at high pH, which is ascribed to the electrostatic repulsion of SL by the membrane. Further improvements in the ratio of retention of SL and lactose were achieved by changing the operational pressure. The best suited membrane was used in a final 10-rounds diafiltration, which enabled total separation of SL and lactose. The study also reveals that while charge differences between solutes can be utilized during nanofiltration, the trade-off between electrostatic repulsion and pore swelling must be addressed when optimizing the nanofiltration process.

1. Introduction

Some of the most bioactive components of human breast milk are the so-called 'human milk oligosaccharides' (HMOs). HMOs are believed to have a great impact on the development of infants but are, however, absent in all currently commercially available infant formulas [1]. Commercially sustainable processes for production of HMOs are therefore of great interest for both society and the dairy industry. 3′-sialyllactose (SL) is an HMO molecule which is currently produced from lactose and casein glyco macro peptide (cGMP) via an enzymatic reaction that relies on the presence of a large excess of lactose [2–4]. Once produced, the separation and purification of SL from lactose is a crucial step, as the SL added to the commercial formula should be as pure as possible. Traditionally, when the production of SL has been performed at small scale, separation of lactose and SL has been achieved by anion exchange chromatography [5,4,3,2]. However, considering the economy of a large scale production process and the fact that anion exchange chromatography relies on the addition of buffers and eluents – which would then need to be removed before the SL can be considered for consumption – an alternative process for the efficient purification of SL from lactose is necessary for the addition of SL to infant formula to become a reality.

Since SL has a molecular size that is almost double than that of lactose (with molecular weights of 632.5 g/mol and 342.3 g/mol respectively) (Fig. 1), the separation of SL and lactose by nanofiltration (NF) should be a possibility and is investigated in this work. For the effective separation of SL and lactose, a NF membrane must demonstrate high level of SL retention and allow lactose to pass through the membrane, resulting in a low level of lactose retention. NF has previously been successfully applied for the separation

* Corresponding author.
E-mail address: mp@kt.dtu.dk (M. Pinelo).

http://dx.doi.org/10.1016/j.seppur.2014.10.012
1383-5866/

of small molecules with size ratios similar to lactose and SL [6]. Whilst the size of the membrane pores creates a physical barrier, the interaction between the charges of the membrane material and the solute creates an electrostatic barrier. Retention of a molecule usually depends on a complex combination of the physical and electrostatic barriers, therefore it is hard to predict the exact behaviour of a separation and empirical data obtained in systematically conducted experiments becomes an indispensable prerequisite for the optimization of a NF process.

Despite the fact that lactose has commonly been the subject of NF, in most cases the aim was to retain lactose while letting e.g. salts pass through the membrane. In other studies, lactose was a part of complex mixtures of oligosaccharides and the ratios between the retentions of glucose, lactose and specific oligosaccharides were manipulated by altering the filtration conditions and membrane materials [7]. The effect of changing pH during NF of charged saccharides has been thoroughly reviewed [8,9]. In general, an increased pH will increase the charge density, of the membrane, and thereby increase the retention of negatively charged molecules via repulsion. The effect of increased pH has been studied extensively for NF involving inorganic salts, and while literature concerning NF separation involving organic acids at increased pH is scarcer, the same pattern has been demonstrated [8,10,11].

In this study, the suitability of NF for the separation of SL from lactose was evaluated. The basic hypothesis behind the study was that more of the charged SL molecules (pKa values for SL and lactose are 2.6 and 11.98 respectively) should be retained at high pH, while retention of lactose should be unaffected by the membrane charge (all other things being equal). To test this hypothesis, filtration experiments were conducted at pH ranging from 4 to 10. Albeit restricted by availability, a selection of membranes with varying material and pore sizes were evaluated to establish the effectiveness of NF as a method for the separation of SL and lactose. Since the final goal of this study was to develop a method capable of purifying SL – along with an understanding of the mechanisms involved in the separation – a final diafiltration step, which consisted of consecutive rounds of filtration, was performed.

2. Materials and methods

2.1. Membranes

Four NF membranes (ETNA01PP, NP010, NTR7450 and NF45) were used in this study. The main characteristics are summarized in Table 1. Where no citation has been indicated, the values are based on manufacturers' information. All the used membranes were washed in ethanol for 10 s and rinsed in Millipore water before use and the initial water permeability was measured (Table 1).

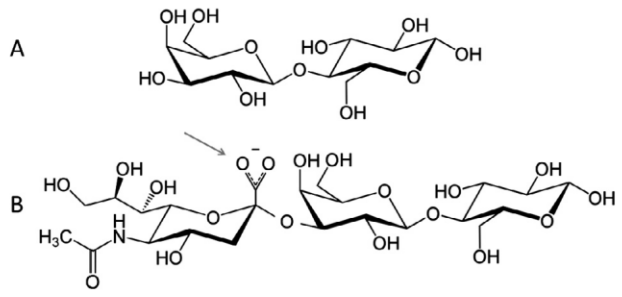


Fig. 1. The structures of (A) β-lactose, and (B) 3′-sialyllactose. The position of the negative charge of SL is indicated by an arrow.

2.2. Chemicals

β-Lactose (30% α-anomer, ⩾99% total lactose basis) from Sigma–Aldrich (Steinheim, Germany) and SL (98% pure) from Carbosynth (Compton, United Kingdom) at different concentrations were used to prepare the model solutions (Table 2). Casein glycomacropeptide (cGMP) in the form of the commercially available product Lacprodan® cGMP-20 was a gift from Arla Foods amba (Viby, Denmark). It contained 5.7% (w/w) of covalently linked sialic acid and was used as the donor molecule for enzymatic production of 3′-sialyllated-lactose (with lactose as acceptor) using an engineered sialidase, Tr13, from *Trypanosoma rangeli* reacting for 60 min at 30 °C and pH 6.5. The Tr13 enzyme was produced in *Pichia pastoris* and purified as described previously [2]. The enzyme and cGMP residues were removed from the post reaction mixture by ultrafiltration using a 10 kDa commercially available Alfa Laval RC70PP membrane before the sample was frozen and stored until it was used for filtration experiments.

2.3. Filtration unit

All experiments were performed at room temperature. A new membrane was used for each experiment except when the water permeability of the used membrane could be fully recovered. All filtration experiments were conducted in a magnetically stirred dead-end cell (Amicon 8050, Millipore, USA) with a working volume of 50 mL and an effective membrane area of 13.4 cm^2. Once inside the cell, the membranes were pre-pressurized to 4 bar for at least 30 min. The pressure in the cell was controlled by feeding nitrogen into the cell and the permeate was collected in a beaker placed on an electronic scale to monitor the permeate flux. The sample volume in all cases was 10 ml and the filtration was stopped when the permeate volume reached 5 ml. It is clear that for a well retained solute (i.e. SL) the concentration dependent Donnan exclusion will change significantly during the experiment. In all experiments, the permeate was (after discarding the first 0.5 ml) collected and frozen for later composition analysis. For single run filtration experiments the membranes were washed after the filtration before the water permeability was measured again. The operation time for each experiment varied due to the difference in flux, which could again affect the compaction behavior for the different membranes. However this effect should be negligible since the membranes were pre-pressurized and since the experiments were run at very low pressures compared to nanofiltration experiments in general.

Diafiltration experiments were carried out as a series of 10 consecutive 10 ml Amicon filtrations with a volume reduction of 50% from 10 ml to 5 ml. After each round of filtration, the retentate remained in the cell and the volume was readjusted to 10 ml by the addition of Millipore water.

2.4. Analysis of permeate and retentate samples

Concentrations of 3′-sialyllactose and lactose in the permeate, retentate and feed solutions were measured by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) analysis using a CarboPac™ PA100 (4 mm × 250 mm) analytical column equipped with a CarboPac™ PA100 (4 mm × 50 mm) guard column (Dionex Corp., Sunnyvale, CA) on a Dionex ICS-3000 system (Dionex Corp., Sunnyvale, CA). The operating conditions and analysis procedure have been previously described by Zeuner et al. [4].

2.5. Calculated parameters

Permeate flux (J) was calculated by:

Table 1
Specifications of the four membranes used in this study.

Membrane Top layer composition	ETNA01PP PVDF	NP010 PES	NTR-7450 SPES	NF45 PA
MWCO (Da)	1000	1000–1400	600–800	200–400
Max. temperature (°C)	60	95 [12]	60 [13]	45 [13]
Max. pressure (bar)	10	40	40	41
pH range	1–11	0–14	2–11	3–10
Zeta potential at pH 7 (mV)	n.a.	−12 [14]	−17 [15]	−12.5 [16]
Isoelectric point (pH)	n.a.	App. 3 [14]	3–4 [15]	6.5 [17]
Contact angle (°)	n.a.	28.5 [14]	69.6 ± 5.9 [15]	40 ± 5.2 [17]
Initial water permeability (1 bar $^{-1}$ m $^{-2}$ h $^{-1}$)	30.5	19.3	11.3	4.1

PVDF: polyvinylidene fluoride; PES: polyethersulphone; PA: polyamide; SPES: sulfonated PES.

Table 2
Specifications of the solutions used in this study.

Solution		Model A	Model B	Authentic
3′-Sialyllactose	mM	0.5[a]	0.5[a]	6[b]
Lactose	mM	0.5[a]	100[a]	>500[b]

[a] Prepared from shelf Chemicals.
[b] Enzymatically produced 3′-siallyllactose, quantified by HPLC.

$$J = \frac{1}{A}\frac{dV_p}{dt} \quad (1)$$

where A is the effective membrane area (m^2), V_p is the permeate volume (L) and t is the filtration time (h).

Water permeability (L_p) was acquired from:

$$L_p = \frac{J_w}{TMP} \quad (2)$$

where J_w is water permeate flux (L m^{-2} h^{-1}) and TMP is transmembrane pressure (bar).

The decline in water permeability was obtained from a comparison of the membrane water permeability (L_p) before and after the filtration:

$$\textit{Decline in water permeability}\ (\%) = \frac{L_{pi} - L_{pf}}{L_{pi}} \times 100 \quad (3)$$

where L_{pi} and L_{pf} are the water permeability of the new and used membranes (before and after filtrations), respectively.

Average observed retention (R_{obs}) of solutes was defined as:

$$R_{obs}\ (\%) = \left(1 - \frac{Cp}{(C_R + C_F)/2}\right) \times 100 \quad (4)$$

where C_p is the solute concentration in the permeate, C_R is the solute concentration in the retentate and C_F is the solute concentration in the feed solution.

Recovery was calculated as:

$$\textit{Recovery}\ (\%) = \frac{C_R \times V_R}{C_0 \times V_0} \times 100 \quad (5)$$

where C_R and V_R are the solute concentration and volume of the retentate and C_0 and V_0 are the initial solute concentration and volume. Purity of SL was defined as:

$$\textit{Purity of SL}\ (\%) = \frac{C_{R,SL}}{C_{R,SL} + C_{R,Lac}} \times 100 \quad (6)$$

where $C_{R,SL}$ and $C_{R,Lac}$ are the respective concentrations of SL and lactose in the retentate.

To model the diafiltration process, a formula describing the retentate concentration of a solute as a function of the retentate–permeate ratio and rounds of filtration was deducted as follows:

$$n_F = n_R + n_P \Rightarrow C_F V_F = C_R V_R + C_P V_P$$

where n, C, and V defines the solute amounts, concentrations and volumes respectively in feed solution, retentate, and permeate.

Therefore

$$C_F = ((1 - N)C_R + NC_P)$$

where the ratio $N = \frac{V_P}{V_F} \Rightarrow V_P = NV_F$ and $V_R = (1 - N)V_F$.

From Eq. (4) it is clear that

$$C_P = \frac{(1 - R)(C_F + C_R)}{2}$$

So

$$C_F = (1 - N)C_R + N\frac{(1 - R)(C_F + C_R)}{2} \Rightarrow C_R = C_F\frac{2 - (N(1 + R))}{2 - (N(1 - R))}$$

$$C_{R,n} = C_{R,n-1}\frac{2 - (N(1 + R))}{2 - (N(1 - R))} \quad (7)$$

where $N = V_P{:}V_F$, $C_{R,n}$ is the concentration of a solute in retentate after n rounds of filtration and $C_{R,n-1}$ is the solute concentration in the retentate after the previous round. It should be noted that $C_{R,0} = C_{F,0}$.

In the diafiltration experiments carried out in this study, the volume reduction was 50%, and based on Eq. (4) and mass balance, it therefore follows that:

$$C_{R,n} = C_{R,n-1}\frac{3 + R_{obs}}{6 - 2R_{obs}} \quad (8)$$

where $C_{R,n}$ is the concentration of a solute in retentate after n rounds of filtration and $C_{R,n-1}$ is the solute concentration in the retentate after $n-1$ rounds. It should be noted that $C_{R,0}$ is double the initial feed concentration in the original volume in this equation.

Experimental diafiltration recovery and purity was based on C_R, calculated as:

$$C_{R,n} = C_{R,n+1} + C_{P,n+1} \quad (9)$$

where $C_{R,n}$ is the concentration of a solute in retentate after n rounds of filtration and $C_{P,n+1}$ and $C_{R,n+1}$ are the solute concentrations in the retentate and permeate after the next round of filtration. It should be noted that $C_{R,10}$ and $C_{P,1..10}$ are all experimental data.

3. Results and discussion

3.1. Effect of pH on permeate flux and retention

The permeate fluxes observed for the four tested membranes were in good correlation with the molecular weight cut-off (MWCO) of each of them, regardless of the pH (Fig. 2) as well as other factors that could affect permeability e.g. ratio of porosity over tortuosity or skin layer thickness. ETNA01PP provided the highest permeate flux, which was approximately 10 times higher

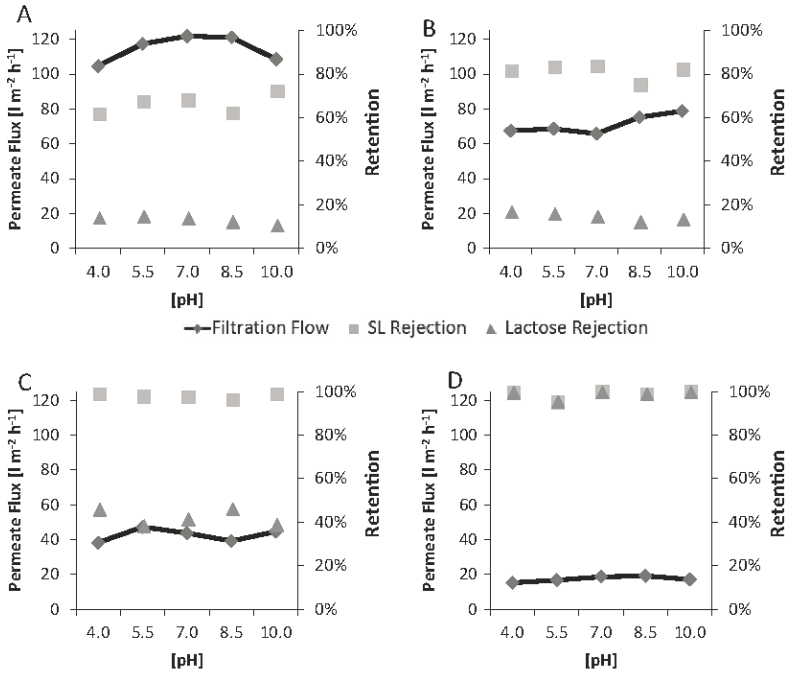


Fig. 2. The performances of the four membranes (A) DSS-ETNA01PP, (B) NP010, (C) NTR-7450 and (D) NF-45, at 5 different pH (ranging from 4 to 10). All data shown in the diagrams was obtained from filtration experiments using a mixture of lactose and SL, both at a concentration of 0.5 mM and an applied pressure of 4 bars. The permeate flux is given on the left *y*-axis [l m $^{-2}$ h $^{-1}$] whereas the retention data for lactose and SL is given on the right *y*-axis [%]. The data shown in the figure represents an average coefficient of variation <10%.

than the one provided by NF45, whose pore size (MWCO) was also lower (Table 1). No significant variations in the permeate flux were immediately detected when the experiments were performed at different pH – from pH 4 to pH 10-, however a decrease in permeability was observed after filtration for at least one of the membranes (NP010), as will be discussed later (part 3.2). When this drop in water permeability for the NP010 membrane is taken into account, it could be speculated that the slight increase in flux with pH, which at first hand identifies as an artifact, could in fact be a result of a less fouled membrane. The MWCO of the membrane had a clear influence on the retention of lactose and SL. Although the MWCO of ETNA01PP would theoretically let both lactose and SL pass through the membrane, there was an observable difference in retention of the two molecules (~80% retention of SL versus ~18% retention for lactose). While a difference in retention can be to some extent ascribable to the difference of molecular weight of the solutes, to attain a difference in retention of the observed magnitude, it seems likely that the electrostatic repulsion between the membrane surface and SL played also a role in this particular separation. The role of electrostatic repulsion on the relative retentions of SL and lactose is confirmed by a comparison to previous studies with similar NF membranes, where during filtration of uncharged saccharides with size ratios similar to that of SL and lactose where the observed differences in retention were much smaller [7]. Both the NP010 and the NTR-7450 membranes showed promising separation performances, since the retention of SL was high (80% for NP010 and 100% for NTR-7450) and the lactose retention was much lower (20% and 60%, respectively). Since the skin layer of both membranes is made of PES, the better performance of the NTR-7450 as compared to NP010 can be to some extent explained by a lower MWCO (Table 1). However, it is clear that there is electrostatic repulsion between the NP010 membrane and SL, as the pore size (MWCO) of the membrane should easily allow SL to pass through. Both the electrostatic repulsion and the smaller pore size contribute to the better performance of NTR-7450 which may be even more charged than the NP010 due to the its content of sulfonated PES.

Since electrostatic repulsion plays a role in the higher retention of SL, it would be expected that pH would also play a major role in the level of SL retention. A higher pH should promote a higher charge density on the membrane that would repel the charged acid, which should result in a higher level of SL retention. The fact that the increased pHs did not have an effect on the level of SL retention could be presumed to be due to swelling of the membranes, which is a phenomenon generally accepted in the literature to occur for NF membranes at high pH [8]. Although the mechanism of membrane swelling is unclear, previous studies have argued that the increased concentration of counter ions inside the pores at high pH (sodium in this case) could cause pore swelling and the charge density of the pore walls could force open the pores by simple electrostatic repulsion [7,18–20]. It has been shown that this suggested pore swelling mechanism in nanofiltration membranes is highly dependent on the nature of the specific counter ion adsorbing to the membrane [20]. Another mechanism that has been suggested to explain NF membrane pore swelling at high pH is that the layer of adsorbed water on the pore wall is packed closer around the counter ions at high pH, making the pore wider as a result [21]. Despite the small changes in the actual values of SL retention at high pH, it could be speculated that the decrease in retention of SL occurring at pH 8.5 for the membranes with the larger pore size (ETNA01PP and NP010) could be due to an initial reduction in retention due to the pore swelling, followed by an increase in retention at pH 10 where the electrostatic repulsion takes dominance. This is further supported by the decrease in permeate flux for the ETNA01PP membrane observed at pH 10, presumably caused by the increase in osmotic pressure resulting from the increased retention For the tighter membrane NF45, no effect on rejection is seen when changing the pH. This is in accordance with previous findings by Dalwani et al. [22], who shows that for a similar tight polyamide membrane changes in pore size

are only observed when pH exceeds 10. A drop in lactose rejection as a result of the likely pore swelling was not detected since lactose retention levels were already very low (>20%).

3.2. Effect of pH on water permeability decline

Although changes in pH during filtration did not have a great net effect on the solute retentions, it did cause a dramatic decline in the water permeability of some of the membranes (Fig. 3). The most prominent effect of the pH increase was observed for the NP010 membrane, where a decline in water permeability was pronounced at low pH (40% at pH 4), whilst almost no effect was observed at pH 8 and above. At high pH, a significant decrease in the water permeability was only observed for the NTR-7450 membrane (of approximately 10%), whilst at low pH a water decline was observed for the DSS-ETNA01PP and the NP010 membrane. The membranes in which an effect of pH on the water permeability decline was evident were the ones with larger pore sizes, as shown by the membrane properties outlined in Table 1. It could be hypothesized that in the two cases where pH affected water permeability (ETNA01PP and NP010), the chance that SL would be adsorbed or entrapped within the pores of the membrane decreases at higher pH –where electrostatic repulsion between SL and the membrane is likely to be stronger. In this way, pore blocking is avoided, and the initial water permeability of the membrane can be more easily restored. Since the MWCOs of NF45 and NTR-7450 are so small, pore blocking in these cases are unlikely and thusly the decline in water permeability should most likely be ascribed to the compaction effect. Based on the water permeability results and the high ratio of retention between SL and lactose, and considering the high retention of SL, both NTR-7450 and NP010 were selected as candidates for future diafiltration.

3.3. Effect of pressure on permeate flux and retention

When the pressure in the cell was decreased, the competitive advantage of smaller molecules, such as water, over the larger molecules (the solutes) to pass through the membrane decreased, which resulted in an increased transport of solutes across the membrane by convection and a reduction of concentration polarization close to the membrane. Although the flux of the molecules (including the solutes) in general is expected to increase with increased pressure, the competitive advantage of water over solutes will increase reducing the levels of solute transported across the membrane by convection and thusly in effect increase the retentions of SL and lactose. Also, when subjected to an increased pressure, a membrane will be compressed; and as a consequence, the pore size may be reduced (sieving effect) [6]. Thusly in an effort to reduce the lactose retention of the NTR-7540 membrane, filtration experiments were carried out at varying lower pressures (Fig. 4). Although increasing the pressure and thereby the SL retention of the NP010 membrane may seem enticing, the increased pressure would certainly influence the lactose retention much more than the SL retention, as sieving effects are predominant for neutral solutes [6,23].

Reducing the pressure did, in fact, reduce the retention of lactose as expected. However, the magnitude of this reduction was rather large (38%) compared to most previous studies. Data sets from studies by Xu et al. [24,25] showed equal performances, with a large reduction in retention at reduced pressures, but it can be argued that in general the effect of pressure on rejection is most significant at low pressures. Interestingly, the SL retention was not affected by the reduced pressure, an indication that the (very

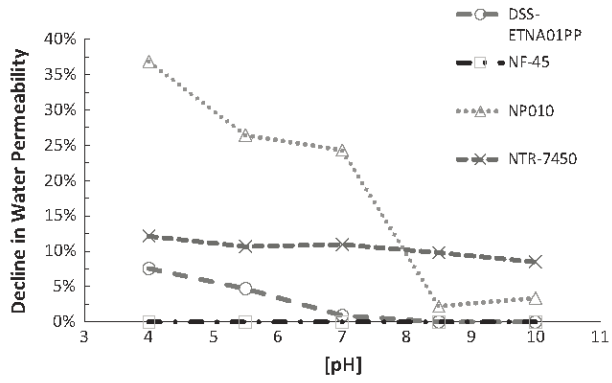


Fig. 3. To illustrate the membrane recovery of the four membranes at different pH, the decline in water permeability has been plotted against the pH. The data in the plot illustrates the water permeability measured before and after the same experiments from which the data in Fig. 2 have been obtained.

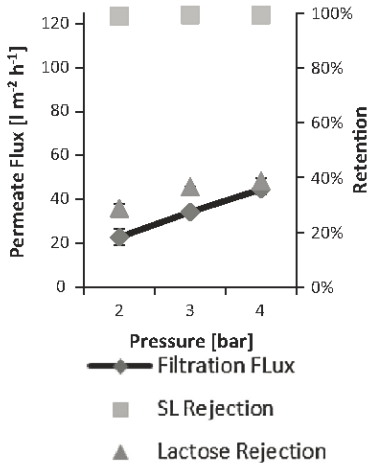


Fig. 4. Effect of pressure on permeate flux and retention for the NTR7540 membrane. All experiments were conducted at pH 10, filtering a mixture of SL and lactose both at a concentration of 0.5 mM. The permeate Flux is given on the left *y*-axis [l m $^{-2}$ h $^{-1}$] whereas the retention data for lactose and SL are given on the right *y*-axis [%].

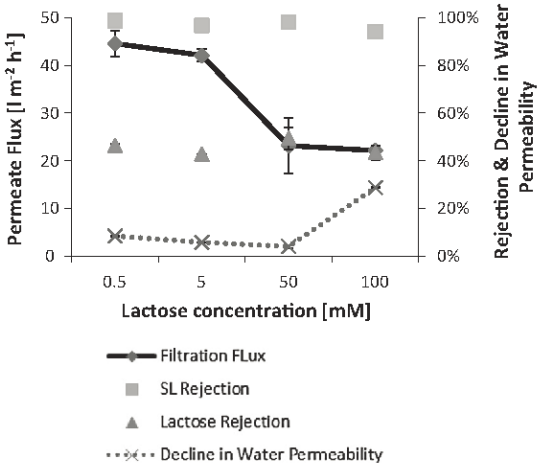


Fig. 5. The effect of higher lactose concentrations on the filtration performance of the remaining candidate membrane (NTR-7450) was investigated and the results are given in the figure. The permeate flux is given on the left *y*-axis [l m $^{-2}$ h $^{-1}$] whereas the retention data for lactose and SL is given on the right *y*-axis [%]. Besides the varying concentrations of lactose, the concentration of SL was kept constant at 0.5 mM for all the experiments. The applied pressure and pH was kept constant at 4 bar and pH 10 for all the experiments.

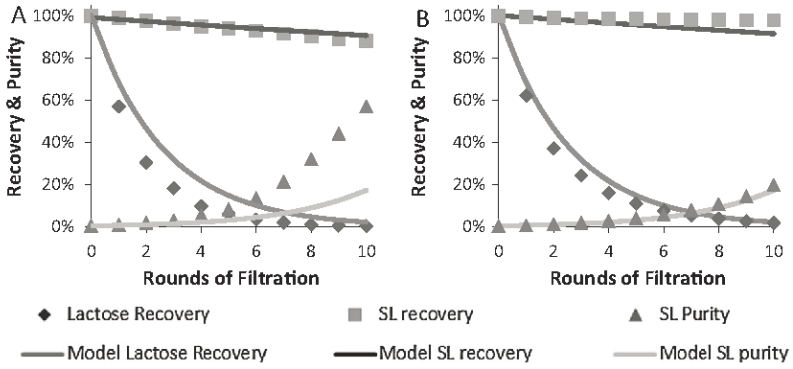


Fig. 6. During 10 rounds of diafiltration, permeate was collected while the retentate volume was reduced to 50% of the starting volume. Based on permeate concentrations of lactose and SL, the corresponding retentate concentrations were calculated and SL purity as well at SL and Lactose recovery were plotted (based on Eqs. (5), (6) and (9)). The feed solutions for the experiments were (A) a model solution of 100 mM lactose and 0.5 mM SL, and (B) a mixture of SL and lactose obtained from the enzymatic production of SL. All the filtrations were carried out a constant pressure of 4 bar. The results were plotted next to the model based on observed retention as input parameters 46.3% and 98.8% retention of lactose and SL respectively.

limited) transport of SL across the membrane is purely due to diffusion rather than convection (and – as it has been seen throughout the study – under strong influence of the electrostatic repulsion to the membrane).

3.4. Effect of lactose concentration

In the mixture obtained from enzymatic production of SL, the lactose concentration is much higher than the SL concentration, and therefore a series of experiments in which lactose concentration was varied was performed (Fig. 5). As it can be seen, the ratio between SL and lactose retentions was unaffected by increasing the lactose concentration. An expected, a decrease in permeate flux was observed for high lactose concentrations. Previous works have also reported significant decreases in permeate flux for increased concentrations of lactose, although the purpose of most of the literature reporting the NF of lactose is to increase the retention of lactose in order to separate it from monosaccharides [26,27]. In one of these studies, a mixture of galacto-oligosaccharides, lactose and monosaccharides were fractionated by NF and it was concluded that the flux decrease caused by increased lactose concentration was partially due to the osmotic pressure resulting from higher feed concentrations, which decrease the primary driving force, pressure, during the operation [7]. It can be seen from Fig. 5 that an increase in lactose concentration was not accompanied by a decline in water permeability at lactose concentrations of 50 mM or less, which contradicts the decline in water permeability reported in 3.2. The smaller decline in water permeability in this experiment is ascribed to the variance in the individual membranes (cut outs) used in the experiments. However, at a lactose concentration of 100 mM the water permeability started to decline, which was probably ascribable to the effect of fouling e.g. adsorption or pore blocking, whose effect was higher at high lactose concentrations. A previous study on separation of whey proteins and lactose, by using polyethersulfone membranes, concluded that lactose could not be adsorbed to the membrane and thus contribute to fouling, although the lactose retention was highly influenced by the occurrence of other species in the mixture [28]. Since the possibility to reuse the membrane is always desirable, particularly in large industrial processes, one should consider not surpassing a lactose concentration of 50 mM lactose in order for the membrane to show high performance for an extended time, although in an industrial setting the problem might solve itself as frequent cleaning of the membrane is a requirement in food processing.

3.5. Diafiltration

A lactose concentration of 100 mM was selected as the starting concentration of the model solution for the diafiltration experiment despite the resulting decline in water permeability because the ratio SL: lactose was the same as the enzymatically produced solution. The results obtained from the model (Eq. (8)) fitted very well those obtained from the model solution and the enzymatically obtained solution (Fig. 6A and B). Surprisingly, the diafiltration experiment performed better than expected in terms of purity of SL, which increased much faster than predicted by the model.

The rapid increase in SL purity was ascribable to the lactose concentration declining more rapidly than expected, which was caused by lower lactose retention in the real diafiltration than predicted by the model. The same rapid decline of lactose recovery was observed for the authentic solution. When comparing the diafiltration of the model solution to the diafiltration of the authentic solution (Fig. 6B vs. A), it was noticed that the increase in SL purity for the authentic solution did not increase as quickly but followed the model prediction. However this could be explained by the concentration of the model solution being 10 times lower than the authentic solution. Based on the lactose recovery, it would be expected that the SL purity would increase quicker than predicted by the model, although it would take several filtration rounds before the difference would be noticeable. An extrapolation of the values of both diafiltration experiments suggests that a purity of >99% could be achieved if ten more purification rounds were performed. Based on the Eq. (7), it seems clear that a diafiltration involving a volume reduction of more than 50% of the initial volume would reduce the number of rounds required to attain the same values of SL recovery and purity. However, since the filtration flux is influenced by the lactose concentration, further experiments would be needed in order to find a balance between a larger volume reduction and a lower number of rounds.

4. Conclusion

The successful separation of SL and lactose using a PES NF membrane (NTR-7450) with a MWCO of 600–800 Da and a zeta potential of −17 (at pH7) was achieved. The observed high retention of SL (close to 100%) was concluded to be the result of both physical and electrostatic retention by the membrane. Although it was not possible to reduce the retention of lactose by manipulation of the filtration conditions, it was concluded that the NTR-7450 membrane will be suitable for purification of SL in a large scale diafiltra-

tion set up. It was clear that changing the pH did not have a net effect on the retention of SL for any of the examined membranes, probably (with exception to the NF45 membrane) due to a tradeoff between charge repulsion and pore swelling induced by high pH. Although not in within the scope of this study, an effort to shed light on the exact mechanisms responsible for the observations at high pH could be of great value to the field of NF. If implemented as the last step of an integrated membrane system as intended, it would be of interest to investigate if the same positive results can be achieved in a continuous diafiltration setup. Finally, the results of this study suggest that NF with negatively charged membranes should always be considered for purification of reaction products and other compounds of interest that are traditionally purified by anion exchange chromatography, when the remaining conditions allow for it.

Acknowledgements

We acknowledge the Danish Strategic Research Council (DSF) for financial support to the strategic project "Enzymatic production of human milk oligosaccharides". Arla Foods amba is acknowledged for the CGMP gift.

References

[1] L. Bode, Human milk oligosaccharides: every baby needs a sugar mama, Glycobiology 22 (2012) 1147–1162, http://dx.doi.org/10.1093/glycob/cws074.
[2] C. Jers, M. Michalak, D.M. Larsen, K.P. Kepp, H. Li, Y. Guo, et al., Rational design of a new *Trypanosoma rangeli* trans-sialidase for efficient sialylation of glycans, PLoS One 9 (2014) e83902, http://dx.doi.org/10.1371/journal.pone.0083902.
[3] M. Michalak, D.M. Larsen, C. Jers, J.R.M. Almeida, M. Willer, H. Li, et al., Biocatalytic production of 3′-sialyllactose by use of a modified sialidase with superior trans-sialidase activity, Process Biochem. 49 (2014) 265–270, http://dx.doi.org/10.1016/j.procbio.2013.10.023.
[4] B. Zeuner, J. Luo, C. Nyffenegger, V. Aumala, J.D. Mikkelsen, A.S. Meyer, Optimizing the biocatalytic productivity of an engineered sialidase from *Trypanosoma rangeli* for 3′-sialyllactose production, Enzyme Microbiol. Technol. 55 (2014) 85–93, http://dx.doi.org/10.1016/j.enzmictec.2013.12.009.
[5] J. Holck, D.M. Larsen, M. Michalak, H. Li, L. Kjærulff, F. Kirpekar, et al., Enzyme catalysed production of sialyllated human milk oligosaccharides and galactooligosaccharides by Trypanosoma cruzi trans-sialidase, N. Biotechnol. 31 (2014) 156–165, http://dx.doi.org/10.1016/j.nbt.2013.11.006.
[6] A.K. Goulas, P.G. Kapasakalidis, H.R. Sinclair, R.A. Rastall, A.S. Grandison, Purification of oligosaccharides by nanofiltration, J. Membr. Sci. 209 (2002) 321–335, http://dx.doi.org/10.1016/S0376-7388(02)00362-9.
[7] Y.M. Feng, X.L. Chang, W.H. Wang, R.Y. Ma, Separation of galacto-oligosaccharides mixture by nanofiltration, J. Taiwan Inst. Chem. Eng. 40 (2009) 326–332, http://dx.doi.org/10.1016/j.jtice.2008.12.003.
[8] J. Luo, Y. Wan, Effects of pH and salt on nanofiltration—a critical review, J. Membr. Sci. 438 (2013) 18–28, http://dx.doi.org/10.1016/j.memsci.2013.03.029.
[9] J. Tanninen, Importance of Charge in Nanofiltration, 2004. <http://books.google.dk/books/about/Importance of Charge in Nanofiltration.html?id=nlt8AAAACAAJ&pgis=1> (accessed June 19, 2014).
[10] A. Morao, A. Alves, M. Afonso, Concentration of clavulanic acid broths: influence of the membrane surface charge density on NF operation, J. Membr. Sci. 281 (2006) 417–428, http://dx.doi.org/10.1016/j.memsci.2006.04.010.
[11] V. Freger, Separation of concentrated organic/inorganic salt mixtures by nanofiltration, J. Membr. Sci. 178 (2000) 185–193, http://dx.doi.org/10.1016/S0376-7388(00)00516-0.
[12] R. Hepsen, Y. Kaya, Optimization of membrane fouling using experimental design: an example from dairy wastewater treatment, Ind. Eng. Chem. Res. 51 (2012) 16074–16084, http://dx.doi.org/10.1021/ie302450s.
[13] B. Van der Bruggen, Nanofiltration as a treatment method for the removal of pesticides from ground waters, Water Supply 17 (1999) 55–63.
[14] C. Aydiner, Y. Kaya, Z. Beril Gönder, I. Vergili, Evaluation of membrane fouling and flux decline related with mass transport in nanofiltration of tartrazine solution, J. Chem. Technol. Biotechnol. 85 (2010) 1229–1240, http://dx.doi.org/10.1002/jctb.2422.
[15] K. Boussu, B. Van der Bruggen, A. Volodin, C. Van Haesendonck, J.A. Delcour, P. Van der Meeren, et al., Characterization of commercial nanofiltration membranes and comparison with self-made polyethersulfone membranes, Desalination 191 (2006) 245–253, http://dx.doi.org/10.1016/j.desal.2005.07.025.
[16] J. Cho, Effects of molecular weight cutoff, *f*/*k* ratio (a hydrodynamic condition), and hydrophobic interactions on natural organic matter rejection and fouling in membranes, J. Water Supply Res. Technol. – AQUA. 51 (2002) 109–123.
[17] L. Puro, M. Mänttäri, A. Pihlajamäki, M. Nyström, Characterization of modified nanofiltration membranes by octanoic acid permeation and FTIR analysis, Chem. Eng. Res. Des. 84 (2006) 87–96, http://dx.doi.org/10.1205/cherd.04036.
[18] G. Bargeman, J.M. Vollenbroek, J. Straatsma, C.G.P.H. Schroën, R.M. Boom, Nanofiltration of multi-component feeds. Interactions between neutral and charged components and their effect on retention, J. Membr. Sci. 247 (2005) 11–20, http://dx.doi.org/10.1016/j.memsci.2004.05.022.
[19] A. Bouchoux, H. Balmann, F. Lutin, Nanofiltration of glucose and sodium lactate solutions variations of retention between single- and mixed-solute solutions, J. Membr. Sci. 258 (2005) 123–132, http://dx.doi.org/10.1016/j.memsci.2005.03.002.
[20] A. Escoda, P. Fievet, S. Lakard, A. Szymczyk, S. Déon, Influence of salts on the rejection of polyethyleneglycol by an NF organic membrane: pore swelling and salting-out effects, J. Membr. Sci. 347 (2010) 174–182, http://dx.doi.org/10.1016/j.memsci.2009.10.021.
[21] J. Luo, Y. Wan, Effect of highly concentrated salt on retention of organic solutes by nanofiltration polymeric membranes, J. Membr. Sci. 372 (2011) 145–153, http://dx.doi.org/10.1016/j.memsci.2011.01.066.
[22] M. Dalwani, N.E. Benes, G. Bargeman, D. Stamatialis, M. Wessling, A method for characterizing membranes during nanofiltration at extreme pH, J. Membr. Sci. 363 (2010) 188–194, http://dx.doi.org/10.1016/j.memsci.2010.07.025.
[23] P.-Y. Pontalier, A. Ismail, M. Ghoul, Mechanisms for the selective rejection of solutes in nanofiltration membranes, Sep. Purif. Technol. 12 (1997) 175–181, http://dx.doi.org/10.1016/S1383-5866(97)00047-6.
[24] L. Xu, L.-S. Du, J. He, Effects of operating conditions on membrane charge property and nanofiltration, Front. Chem. Sci. Eng. 5 (2011) 492–499, http://dx.doi.org/10.1007/s11705-011-1143-7.
[25] C. Mazzoni, S. Bandini, On nanofiltration Desal-5 DK performances with calcium chloride–water solutions, Sep. Purif. Technol. 52 (2006) 232–240, http://dx.doi.org/10.1016/j.seppur.2006.04.004.
[26] S. Veríssimo, K.-V. Peinemann, J. Bordado, New composite hollow fiber membrane for nanofiltration, Desalination 184 (2005) 1–11, http://dx.doi.org/10.1016/j.desal.2005.03.069.
[27] J. Mmbaga, R.E. Hayes, W. Pikus, A. Afacan, D.C. Bressler, Numerical modelling and characterization of nanofiltration membranes for the separation of carbohydrate mixtures, Can. J. Chem. Eng. 89 (2011) 314–324, http://dx.doi.org/10.1002/cjce.20400.
[28] R.R. de Souza, R. Bergamasco, S.C. da Costa, X. Feng, S.H.B. Faria, M.L. Gimenes, Recovery and purification of lactose from whey, Chem. Eng. Process. Process Intensif. 49 (2010) 1137–1143, http://dx.doi.org/10.1016/j.cep.2010.08.015.

Chapter 6: An integrated membrane system for the biocatalytic production of 3'-sialyllactose from dairy by-products

6.1 Motivation

In this Chapter the advantage of expanding the scope of a PhD project to an overall process, at the interfaces between the different aspects and related technologies is demonstrated. A clear motivator for this study was the upcoming upscaling where handling would undoubtedly become a challenge. An integrated membrane system was not the immediate solution, but was mediated by numerous shortcomings of the pilot plant. A request for a stirred tank reactor for enzyme reaction could not be met and the UF unit used for the pretreatment of the CGMP had an incomprehensive dead-volume which would result in high levels of waste. This led to a realization which in turn was the motivation for the study reported in the following paper: Whereas all steps of 3'SL production in the lab had been done separately at different laboratory benches, by conducting as much of the work as possible in the UF unit almost all of the deficiencies could be turned into advantages. CGMP could be purified in the UF unit and remaining reaction components could be added directly eliminating the handling and dead-volume issues. By circulation of the reaction mixture using the pump on the UF unit without pressure across the membrane the process would be continuously mixed without stirring and the temperature could be maintained by the water cooling system attached. Furthermore this strategy had the benefit of allowing both enzyme and acceptor to pass across the membrane and react with any CGMP which might have adsorbed on the membrane, an improvement compared to the lab scale experiments where donor adsorbed during initial purification was always lost. Finally reaction stop by heat inactivation was redundant since product purification could be started by a turn of a knot, reapplying pressure across the membrane, thereby separating the product (3'SL), and unreacted acceptor (lactose) from CGMP and enzyme.

The main motivation for the study described in this paper was therefore to report the achievement of having designed an integrated membrane system for biocatalytic production of 3'SL, with the added possibility to improve the developed large scale process by identification of a superior UF membrane for the specific purpose.

6.2 Hypotheses and objectives

The objective of this study was to comprehensively communicate the advantages of the integrated membrane system developed during the up-scaling of HMO production carried out in the pilot plant, backed up with sound scientific data.

Thus the main hypothesis to be tested in this study was:

Hypothesis 1: The following observations from large scale cross-flow UF unit will translate well to a laboratory experimental setup using dead-end stirred tank filtration setup:

- Fouling is the main concern with regard to CGMP UF
- Hydrophobic membranes are unsuitable for CGMP filtration due to heavy fouling.
- UF is a viable method for separation of product mixture (3'SL and lactose) from the remaining reaction components.

Hypothesis 2: An additional hypothesis that was tested in this study was that the biocatalytic productivity could be greatly enhanced by application of the integrated membrane system as enzyme could be retained in the EMR and reused accordingly.

In the paper several hypotheses were addressed on the topic of NF as well. However, the previous study deals with this topic in greater detail and is therefore not considered further.

6.3 Experimental considerations

Designing the experiments for this study was a little awkward as part of "the result" was already known. At the time where the scientific experiments for this study were carried out production had been carried out in the pilot plant and NF had been applied for product separation with varying success. Since the NF part of the process had undergone heavy scrutiny in the previous study, only a small series of showcase filtration experiments were set up to report the applicability of the process in its entirety. The main focus on experimental design in this study was concerning the enzymatic membrane reactor. Here identification of a suitable UF membrane was the main focus since the fouling behavior of CGMP had previously caused serious issues in the large scale production. Due to the previous experience with CGMP fouling a suitable 10000 Da regenerated cellulose (RC) UF-membrane had been identified during the development of the pilot plant scale 3'SL production.

However in order to evaluate the various aspects of the proposed membrane system scientifically it was decided to downscale the process again. A disadvantage of the downscaling was that our experimental setup for membrane evaluation relied on the application of dead-end stirred tank filtration units compared to the spiral membrane crossflow filtration unit which was applied in the pilot plant. Thus there was no guarantee that the RC membrane (already identified) would perform as well in the laboratory setting, but this was accepted as a premise because the possibility to identify other suitable membranes (worthy of pilot plant testing) was advantageous. Furthermore it was evaluated that any ideal membrane identified in the stirred tank reactor would be likely perform better, due to reduced fouling, if applied as a spiral membrane in a crossflow setup. Thus it was reasoned that overestimated performance was unlikely to occur with the chosen setup.

6.4 Conclusions

The results from the pilot plant translated well with the laboratory scale experiments as the hydrophilic RC membranes generally outperformed the other types of membranes.

Fouling was a severe influence on membrane performance and the exclusion of charged solutes (Sialic acid and 3'SL) was repelled by the hydrophobic membrane as it was observed in the pilot plant. Additional data supporting that this repulsion was indeed the result of a charged fouling layer (CGMP layer) was also obtained.

A difference between the laboratory and pilot plant experiments were seen with regard to MWCO. It can be concluded that the CGMP during UF behave like a compound around the size of 10000 Da since it (surprisingly) was noted that application of a 10000 Da RC membrane (contrary to the results obtained in the pilot plant) resulted in severe loss of substrate (permeation of CGMP) during pretreatment.

6.5 Paper V

Bioresource Technology 166 (2014) 9–16

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

An integrated membrane system for the biocatalytic production of 3′-sialyllactose from dairy by-products

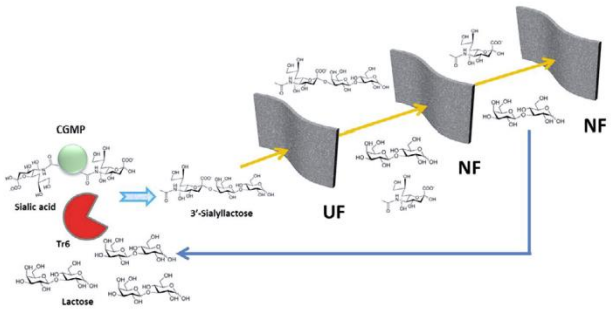
Jianquan Luo [1], Rune T. Nordvang [1], Sofie T. Morthensen, Birgitte Zeuner, Anne S. Meyer, Jørn Dalgaard Mikkelsen, Manuel Pinelo *

Department of Chemical and Biochemical Engineering, Center for BioProcess Engineering, Building 229, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

HIGHLIGHTS

- PLCC regenerated cellulose membrane (5 kDa) can be used to configure the EMR.
- NTR7450 membrane can be used for separation of 3′-sialyllactose and lactose.
- Lactose can be concentrated and recycled by NF45 membrane.
- Tr6 sialidase can be reused in the EMR and retain the activity after centrifugation.
- CGMP residues accumulated in the EMR results in a flux decline and need to be removed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form 30 April 2014
Accepted 2 May 2014
Available online 10 May 2014

Keywords:
Ultrafiltration
Nanofiltration
Enzyme
EMR
Human milk oligosaccharides

ABSTRACT

An integrated membrane system was investigated for the production of 3′-sialyllactose by an engineered sialidase using casein glycomacropeptide (CGMP) and lactose as substrates. CGMP was purified by ultrafiltration (UF) to remove any small molecules present and then an enzymatic membrane reactor (EMR) was used to separate the product and reuse the enzyme. A PLCC regenerated cellulose membrane was found to be the most suitable for both the UF purification and EMR. Subsequently, nanofiltration (NF) was conducted to increase the purity of the 3′-sialyllactose by removing the excess lactose present. The NTR7450 membrane outperformed others in NF due to its high retention of 3′-sialyllactose (98%) and relatively low rejection of lactose (40%). The lactose in the permeate could be concentrated by the NF45 membrane and recycled into the EMR. The described integrated membrane system enables a more economic and efficient enzymatic production of 3′-sialyllactose.

1. Introduction

Human milk oligosaccharides (HMOs) play an important role in the development of a strong immune system in infants by stimulating the growth and activity of beneficial intestinal bacteria (Hickey, 2012). The lack of HMOs in bovine milk requires supplementation of HMOs into infant formula in order to obtain a better simulation of human breast milk (Gopal and Gill, 2000; Urashima et al., 2013). One of the required HMOs is sialyllactose, an HMO model case compound that consists of *N*-acetylneuraminic acid (sialic acid) bound to β-lactose. The scarcity of sialyllactose in nature together with the relatively high cost of commercial sialyllactose (270 euro/g, www.carbosynth.com), encourages the search for efficient approaches for large scale production of sialyllactose.

* Corresponding authors. Tel.: +45 4525 2950; fax: +45 4588 2258.
E-mail addresses: jluo@kt.dtu.dk (J. Luo), mp@kt.dtu.dk (M. Pinelo).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.biortech.2014.05.003
0960-8524/

Sialyllactose can be enzymatically synthesized by trans-sialidases and sialyltransferases (Gilbert et al., 1998; Holck et al., 2014). Tr6 is a sialidase from *Trypanosoma rangeli* that has been engineered to improve its trans-sialidase activity (Paris et al., 2005). Recently, our group reported efficient expression of Tr6 in *Pichia pastoris* after codon-optimization (Michalak et al., 2014). Tr6 can be used to produce 3′-sialyllactose in gram quantities using casein glycomacropeptide (CGMP) and lactose as substrates. Both CGMP and lactose are obtained from side streams in the dairy industry (Michalak et al., 2014; Zeuner et al., 2014). The studies by Michalak et al. (2014) and Zeuner et al. (2014) only focused on production of the mutated enzyme and optimization of the enzymatic reaction. It is well known that the configuration of a suitable product purification technology is a prerequisite for successful manufacturing of enzymatically produced HMOs, especially in large-scale.

Several challenges remain to be solved in order to obtain relatively pure 3′-sialyllactose from the above-mentioned method. First, CGMP, isolated from cheese whey, is a heterogeneous fraction containing casein-derived peptides and varying amounts of sugars including sialic acid, galactose, and *N*-acetylgalactosamine (Abd El-Salam, 2006; Silva-Hernandez et al., 2002). In order to alleviate the burden on downstream separation, the small peptides and salts in CGMP should be removed before feeding enzymes into the reactor. Second, although Tr6 is a sialidase which has been genetically engineered to increase its trans-sialidase activity, its hydrolytic activity has not been fully eliminated and therefore the hydrolysis of sialyllactose inevitably occurs when Tr6 is active (Michalak et al., 2014; Zeuner et al., 2014). In order to reuse the enzyme and avoid extensive product hydrolysis, *in situ* separation of the product from the reactants is necessary. Third, Tr6 can also simultaneously catalyze the release of free sialic acid from CGMP, but the free sialic acid cannot be utilized to synthesize sialyllactose. A high substrate acceptor/donor molar ratio (25:1) can repress the release of sialic acid and maximize the molar yield of sialyllactose (Michalak et al., 2014). However, the excess lactose acceptor needs to be separated from the product in order to ease the burden for the subsequent purification of sialyllactose by chromatography. Moreover, the reuse of acceptors is important, especially when using expensive acceptors such as lacto-N-tetraoses and lacto-N-fucopentaoses (Michalak et al., 2014). Based on the challenges associated with the enzymatic production of sialyllactose using Tr6, an integrated membrane separation system is proposed to purify the sialyllactose product and reuse the enzyme and substrate (Fig. 1). The proposed system consists of four steps of membrane filtration, namely two ultrafiltration (UF) steps and two nanofiltration (NF) steps. CGMP is initially pre-purified by UF1, followed by UF2 when the enzymatic membrane reactor (EMR) is configured. 3′-Sialyllactose and lactose from the EMR permeate can be separated by NF1 and the permeate lactose is concentrated by NF2 and returned into the EMR. Zeuner et al. (2014) found that a 10 kDa regenerated cellulose membrane fully retained the activity of free Tr6 for seven cycles while the transmembrane pressure (TMP) increased due to the accumulation of CGMP in the EMR. Because CGMP can self-assemble and aggregate at high concentrations and low temperatures, refrigerated centrifugation may deposit some CGMP aggregates and allow recovery of the enzyme from the supernatant (see Fig. 1). Although integrated membrane systems have been widely applied in dairy processing and biorefineries (Abels et al., 2013; Luo et al., 2011; Martinez-Ferez et al., 2006; Qi et al., 2012), to the best of our knowledge, this is the first systematic study regarding the application of membrane filtration to the enzymatic production of HMOs.

The present work was undertaken to assess the feasibility of a new integrated membrane system for the production of 3′-sialyllactose by the Tr6 mutant sialidase, using two dairy by-products

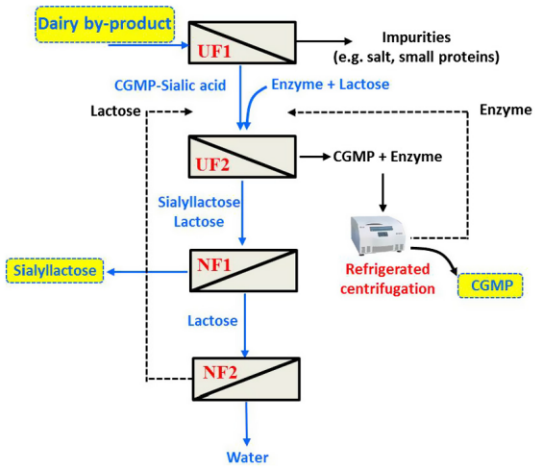


Fig. 1. Schematic diagram of sialyllactose production, purification and enzyme reuse by membrane filtrations and centrifugation. UF1 for purification of CGMP; UF2 for recycling enzyme and removing product; NF1 for separation of sialyllactose and lactose; NF2 for concentration of lactose; refrigerated centrifugation for separation of agglomerate CGMP and free enzyme.

(i.e. CGMP and lactose) as substrates. The objective of the present work was to realize product purification and reuse of both the enzyme and excess lactose (the acceptor substrate), with a focus on the selection of UF and NF membranes with favorable properties in terms of solute rejection, permeate flux, and antifouling performance.

2. Methods

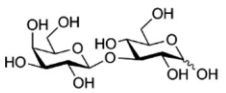
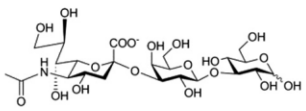
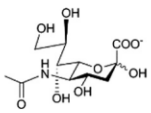
2.1. Chemicals and membranes

Casein glycomacropeptide (CGMP) in the form of the commercially available product Lacprodan® CGMP-20, with a total of 5.7% (w/w) of covalently linked sialic acid was a gift from Arla Foods amba (Viby, Denmark). CGMP is a hygroscopic powder and its moisture content is approximately 17% (w/w). A 3′-sialyllactose standard was purchased from Carbosynth (Compton, United Kingdom). β-Lactose and *N*-acetylneuraminic acid (sialic acid) standards were purchased from Sigma–Aldrich (Steinheim, Germany). The Tr6 enzyme was produced in *P. pastoris* (Michalak et al., 2014) and purified as described previously (Zeuner et al., 2014). The main characteristics of the substrates, products and enzyme are shown in Table 1 (Farías et al., 2010; Gasteiger et al., 2005; Michalak et al., 2014; Neelima et al., 2013; Paris et al., 2005). Seven commercial UF membranes (PLCC, GR81PP, RC70PP, PLGC, GR61PP, Biomax30, and PLTK) and four NF membranes (ETNA01PP, NP010, NTR7450 and NF45) were used in this work, and their main properties are summarized in Table 2 based on the manufacturer's information.

2.2. Experimental set-up and procedure

All the filtration and enzymatic reaction experiments were conducted in a magnetically stirred dead-end cell (Amicon 8050, Millipore, USA). The working volume of the cell was 50 mL and the effective membrane area was 13.4 cm^2. A constant pressure was provided by filling the cell with nitrogen gas while permeate was collected in a tube placed on an electronic scale in order to calculate the permeate flux. All experiments were conducted at a controlled room temperature of 22–24 °C. A new membrane was used for each test except in cases where the permeability of the used membrane could be fully recovered. The substrates for the

Table 1
Main characteristics of substrates, products and enzyme.

Name	Casein glycomacropeptide (CGMP)	β-Lactose	3′-Sialyllactose	Sialic acid (*N*-acetylneuraminic acid)	Tr6 trans-sialidase
Molecular weight (Da)	7500 (Farías et al., 2010)[a]	342.30	633.55	309.27	80,000 (Michalak et al., 2014)
Charge [b]	Negative (pI ~ 4) (Neelima et al., 2013)	Neutral	Negative ($pK_a \sim 2.9$)	Negative ($pK_a \sim 2.6$)	Positive (theoretical pI ~ 8.69 by ProtParam[c])
Structure	Sialic acid 3′-linked to CGMP (Neelima et al., 2013)				PDB: 1WCS (Paris et al., 2005)

[a] Average molecular weight.
[b] Solution pH is approximately 6.5.
[c] Protparam: http://web.expasy.org/protparam/ (Gasteiger et al., 2005); amino acid sequence can be found in Michalak et al. (2014).

Table 2
Main properties of ultrafiltration and nanofiltration membranes.

Membranes		Manufacturer	Molecular weight cut-off (Da)	Surface material
Ultrafiltration	PLCC	Millipore	5000	Regenerated cellulose
	GR81PP	Alfa Laval	10,000	Polyethersulphone
	RC70PP	Alfa Laval	10,000	Regenerated cellulose
	PLGC	Millipore	10,000	Regenerated cellulose
	GR61PP	Alfa Laval	20,000	Polysulphone
	Biomax30	Millipore	30,000	Polyethersulphone
	PLTK	Millipore	30,000	Regenerated cellulose
Nanofiltraztion	ETNA01PP	Alfa Laval	1000	Composite fluoro polymer
	NP010	Microdyn-Nadir	1000	Polyethersulfone
	NTR7450	Nitto-Denko	600–800	Sulfonated polyethersulfone
	NF45	DOW-Filmtec	200	Polypiperazine amide

UF tests were prepared using deionized water while Millipore pure water was used for the NF tests. The pH of the CGMP solution was approximately 6.5.

2.2.1. Purification of CGMP (UF1)

The new polysulphone (PS) and polyethersulphone (PES) UF membranes were first cleaned using a 0.1% NaOH solution for 30 min with a pressure of 1 or 0.1 bar, while the virgin regenerated cellulose (RC) UF membranes were dipped in 5% NaCl solution for 30 min. After cleaning, the water permeability of the membranes was determined in the stirred cell at 1–3 bar. The subsequent UF filtration experiments were carried out in two stages:

2.2.1.1. Rapid evaluation of membrane. A 20 mL CGMP (30 g L^{-1} dry weight, equal to 4 mM assuming its average molecular weight is 7.5 kDa) solution was added to the cell and concentrated at a pressure of 1.5 bar and a stirring speed of 250 rpm. The first 0.5 mL of permeate was discarded and the initial permeate flux was recorded. After another 9.5 mL of permeate was obtained (7.5 mL for the PLCC, GR81PP and GR61PP membranes because their final permeate flux was close to zero), the filtration was stopped and the corresponding retentate was collected for subsequent analysis.

2.2.1.2. Pre-treatment of CGMP for EMR (diafiltration). A 10 mL CGMP (61 g L^{-1} dry weight, 9.3 mM sialic acid bound to CGMP) solution was concentrated to 4 mL with the PLCC and GR81PP membranes at 4 bar, or to 2 mL with the RC70PP membrane at 2.5 bar. Then the retentate was diluted using deionized water to regain the original volume. Three cycles of concentration–dilution (diafiltration) were carried out for the PLCC and GR81PP membranes and five cycles for the RC70PP membrane to ensure that most of the small proteins were washed out.

2.2.2. EMR (UF2)

Enzymatic reactions were performed in two stages:

2.2.2.1. 3′-Sialyllactose production and permeation. After diafiltration, the CGMP retentate was diluted to regain its original volume. A 5 mL CGMP solution was mixed with 5 mL of lactose (234 mM), and 17 μL of purified Tr6 enzyme was added (Zeuner et al., 2014). The reaction was allowed to continue for 30 min (PLCC and GR81PP) or 40 min (RC70PP) without pressure at an agitation of 50 rpm. Afterwards, filtration was conducted at 250 rpm with the same membrane used in diafiltration (UF1) at 4 bar for the PLCC and GR81PP membranes or at 2.5 bar for the RC70PP membrane until 5 mL of permeate was obtained. After filtration the permeate and the retentate tubes were immersed in boiling water for 10 min in order to stop the reaction.

2.2.2.2. Enzyme reuse. Only the PLCC membrane was used for enzyme reuse. A 5 mL diafiltrated CGMP solution was mixed with 5 mL of lactose (234 mM) and 17 μL of purified Tr6 enzyme in the EMR. The reaction was allowed to continue for 20 min without pressure at 50 rpm and then filtration was conducted at 250 rpm and 4 bar until the permeate flux decreased to below a set value. Afterwards, a fresh substrate mixture with a volume equal to the permeate volume just obtained was added into the reactor. The reaction and filtration were carried out for three cycles. After that, the retentate was collected and centrifuged at 4 °C for 15 min (10,000×*g*). The supernatant was returned to the reactor with fresh substrates and another cycle was conducted. The enzyme in the

retentate after the fourth cycle was deactivated by immersing the enzyme containing tube in boiling water for 10 min.

2.2.3. NF separation and concentration

The virgin NF membrane was soaked with 50% ethanol for 30 s before flushing with Millipore water. In order to remove any remaining solvent and to reduce the effect of pressure on membrane performance in subsequent experiments, the membrane was pre-pressurized at 4 bar for 30 min before measuring the water permeability at 1–4 bar. For the NF experiments, a model solution of 0.5 mM 3′-sialyllactose and 0.5 mM β-lactose was used (our preliminary evaluations showed that the effect of 3′-sialyllactose concentration (0.5–4.0 mM) on the rejection was negligible). A 10 mL model solution was added to the cell and a pressure of 4 bar was applied until 5 mL of permeate was obtained (the first 0.5 mL was discarded).

2.3. Analytical methods

Dry weight of CGMP was determined after drying 1 mL of solution at 105 ± 2 °C overnight in an oven. The measurements were repeated three times. Concentrations of 3′-sialyllactose and sialic acid in the permeate and retentate were measured by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) analysis using a CarboPac™ PA100 (4 mm × 250 mm) analytical column equipped with a CarboPac™ PA100 (4 mm × 50 mm) guard column (Dionex Corp., Sunnyvale, CA) on a Dionex ICS-3000 system (Dionex Corp., Sunnyvale, CA). The operating conditions and procedure were described by Zeuner et al. (2014).

2.4. Calculated parameters

Permeate flux (J) was calculated by:

$$J = \frac{1}{A}\frac{dV_p}{dt} \tag{1}$$

where A is the effective membrane area (m^2), V_p is the permeate volume (L) and t is the filtration time (h).

Water permeability (L_p) was acquired from:

$$L_p = \frac{J_w}{TMP} \tag{2}$$

where J_w is water permeate flux ($L\,m^{-2}\,h^{-1}$) and TMP is transmembrane pressure (bar).

Permeability loss was obtained from a comparison of the membrane water permeability (L_p) before and after the filtration:

$$\text{Permeability loss} = \frac{L_{pi} - L_{pf}}{L_{pi}} \times 100 \tag{3}$$

where L_{pi} and L_{pf} are the water permeability of the new and fouled membranes, respectively.

Average observed retention (R_{obs}) of solutes was defined as:

$$R_{obs} = \left(1 - \frac{C_p}{C_{R,av}}\right) \times 100 \tag{4}$$

where C_p is the solute concentration in the permeate and $C_{R,av}$ is the average solute concentration in the retentate during the filtration.

Selectivity was calculated as follows:

$$\text{Selectivity} = \frac{R_{obs,S}}{R_{obs,L}} \tag{5}$$

where $R_{obs,S}$ and $R_{obs,L}$ are the observed retentions of 3′-sialyllactose and lactose, respectively.

3. Results and discussion

3.1. Filtration behavior of CGMP by different UF membranes

A UF membrane with high flux and high retention of CGMP but low level of irreversible fouling is desirable for the filtration of CGMP. Fig. 2a shows that the water flux for the seven studied membranes decreases in the following order: Biomax30 > PLTK > PLGC > GR61PP > RC70PP > GR81PP > PLCC. However, the permeate flux of CGMP solution does not follow the same series, instead it decreases in the following order: PLTK > Biomax30 > PLGC > RC70PP > GR61PP > PLCC > GR81PP. The order of membrane permeate flux could not be directly correlated with pore size (Table 1), probably due to the differences in membrane materials and porosities. In order to clarify the relationship between pore size and membrane permeate flux, the membranes were classified into three groups: PLTK vs. Biomax30, RC70PP vs. GR61PP, and PLCC vs. GR81PP, where the former member of each group had a lower water flux but a higher CGMP permeate flux. With regards to pore size, it can be seen from Table 1 that: PLTK (30 kDa) = Biomax30 (30 kDa); RC70PP (10 kDa) < GR61PP (20 kDa); PLCC (5 kDa) < GR81PP (10 kDa). Within each group, the water flux was lower for the membrane with the lower pore size (except for Biomax30, which may have a very high porosity). However, the permeate flux of CGMP was not only governed by membrane pore sizes but also by the interactions between CGMP and the membrane materials. The surface layers of the PLTK, PLGC, RC70PP and PLCC membranes are made of regenerated cellulose (RC), a highly hydrophilic polymer, which induced a high permeate flux due to its negligible

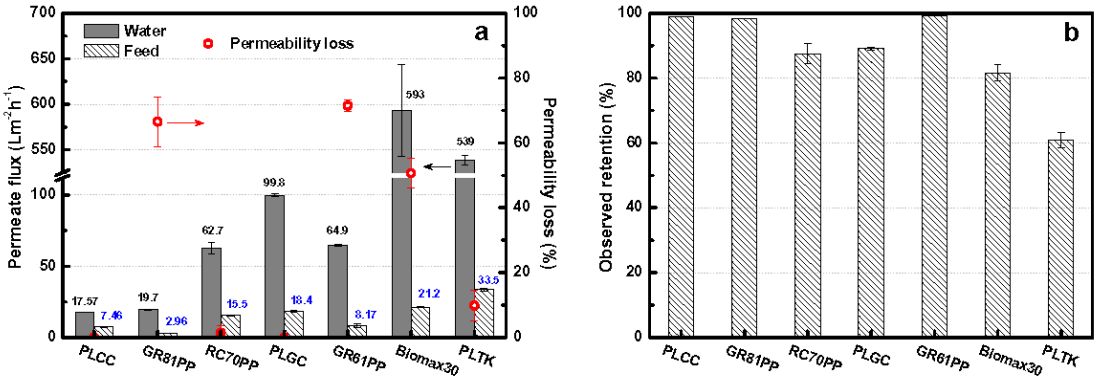


Fig. 2. (a) Permeate flux and permeability loss after filtration, and (b) observed retention of dry weight for different UF membranes. TMP = 1.5 bar; agitation speed = 250 rpm; CGMP concentration = 30 g L^{-1}.

hydrophobic interaction with organic compounds (Giorno et al., 2013; Puro et al., 2010). The permeability loss for the GR81PP, GR61PP and Biomax30 membranes was much higher than for the other membranes, as seen in Fig. 2a, because the PES and PS membranes were easily fouled by proteins (Leberknight et al., 2011). The original permeability could be fully recovered for all the RC membranes except for the PLTK membrane. The loss of permeability observed for the PLTK membrane is probably due to its large pore size, resulting in pore blocking by CGMP. The interaction between CGMP and the membrane materials also affected the degree of retention of CGMP. The CGMP retention was much higher for the PES and PS membranes by comparison to RC membranes with the same pore size (Fig. 2b). The increased CGMP retention of the PES and PS membranes was due to adsorption of CGMP on the surfaces of the membranes.

The negatively charged PLCC (5 kDa) membrane (Chan et al., 2002) could almost fully retain CGMP (~7.5 kDa, negative charge at pH 6.5, see Table 1) by size exclusion and electrostatic repulsion, furthermore, it showed an acceptable permeate flux and no irreversible fouling, rendering it the most suitable membrane for this purpose. The RC70PP membrane with high flux – even if relatively low CGMP rejection – and the GR81PP membrane with high CGMP retention – even if the lowest flux – were also selected for the next study.

In order to remove most small proteins from the CGMP, variable volume diafiltration (VVD) (Foley, 2006) was used as the first step in the integrated membrane system, previously referred to as UF1. As seen in Fig. 3, the average permeate flux increased when the small proteins were washed out of the CGMP, especially for the RC70PP membrane. However, the dry weight loss observed after UF1 with the RC70PP membrane was so high that more than 70% of the CGMP was removed, resulting in a huge loss of substrate, i.e. sialic acid linked to CGMP (Fig. 3b). The dry weight loss was 6% after UF1 with the PLCC membrane and 12% after UF1 with the GR81PP membrane, indicating that the loss of sialic acid combined with CGMP was negligible (Fig. 3b). Although it is time-consuming and may result in a minor loss of the target substrate, the diafiltration operation is required to minimize the presence of impurities in the product (the permeate from the subsequent EMR).

3.2. EMR for production of 3′-sialyllactose

3.2.1. 3′-Sialyllactose production and permeation

After several diafiltration cycles, the Tr6 enzyme and lactose were added to the stirred cell equipped with the same membrane used for the diafiltration (UF1). In the second step of the integrated membrane system (UF2), both high 3′-sialyllactose production and permeation from the EMR are desirable. As shown in Fig. 4a, the average 3′-sialyllactose concentration in the permeate and retentate was approximately 1.5 mM for the EMRs with the PLCC and GR81PP membranes, indicating consistency with the results reported by Zeuner et al. (2014). When the RC70PP membrane was used, the average 3'-sialyllactose concentration in the permeate and retentate was only 0.5 mM, confirming that a lot of sialic acid linked CGMP was washed away during UF1. Therefore, the RC70PP membrane could not be used for UF1 and UF2 even if it had an extremely high permeate flux (Fig. 3a). Moreover,

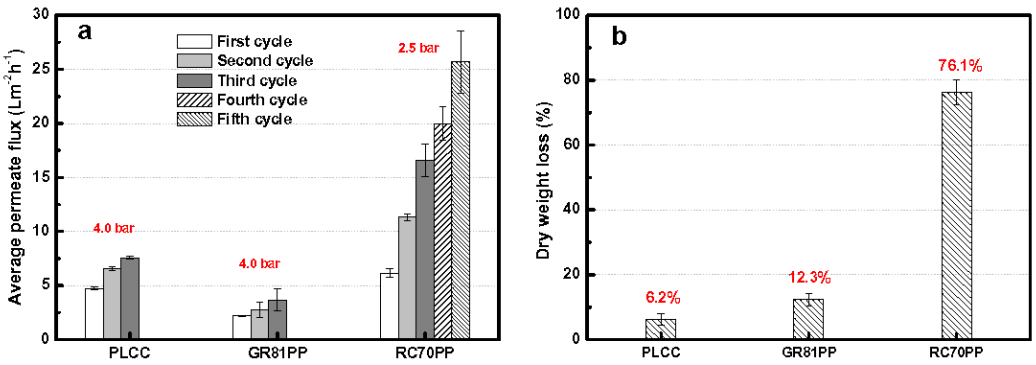


Fig. 3. (a) Average permeate flux and (b) dry weight loss during the CGMP purification operation (UF1) by different membranes. TMP = 4.0 bar for PLCC and GR81PP; TMP = 2.5 bar for RC70PP; agitation speed = 250 rpm; initial CGMP concentration = 61 g L^{-1}.

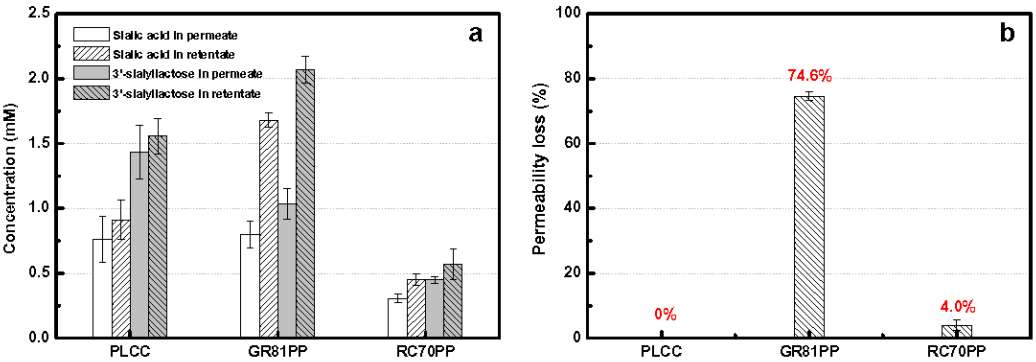


Fig. 4. (a) Product concentrations in permeate and retentate and (b) permeability loss of membranes in enzymatic membrane reactors after integrated UF1 + UF2 process. 5 mL washed CGMP + 5 mL lactose + 17 μL sialidase; agitation speed = 50 rpm during the reaction; agitation speed = 250 rpm during filtration; TMP = 4 bar for PLCC and GR81PP; TMP = 2.5 bar for RC70PP.

0.7–0.8 mM of sialic acid was detected in the permeate after EMR with both the PLCC and GR81PP membranes, implying that the hydrolysis of product (sialyllactose) and/or substrate (sialic acid bound to CGMP) was inevitable in the presence of the Tr6 enzyme.

It can be seen in Fig. 4a that both sialic acid (0.3 kDa) and sialyllactose (0.6 kDa) can freely pass through the PLCC membrane (10 kDa) while the fouled GR81PP membrane (10 kDa) retained higher levels of sialic acid and sialyllactose. The unexpected product retention was probably caused by the CGMP fouling layer in/on the PES membrane, not only narrowing the pore size but also enhancing the electrostatic repulsion (CGMP, sialic acid and sialyllactose are negatively charged; see Table 1). This hypothesis was confirmed by the permeability loss data presented in Fig. 4b, showing that the GR81PP membrane was severely fouled while there was no irreversible fouling layer on the PLCC membrane. Thus, the PLCC membrane was the only choice for the integrated CGMP filtration and EMR process (UF1 + UF2).

3.2.2. Enzyme reuse

In order to reuse the enzyme and reduce product hydrolysis, the product (i.e. 3′-sialyllactose) was removed from the EMR and the fresh substrate mixture was fed in batch. The PLCC membrane was used in the EMR. Fig. 5 shows the flux behavior and product concentration obtained in the permeate with the EMR in batch operation, from which four obvious conclusions can be drawn. First, the permeate flux declined quickly during each cycle due to the increasing CGMP solutes retained on the membrane (higher osmotic pressure and concentration polarization resistance). Second, due to accumulation of CGMP in the EMR upon repetitive addition of new substrate, both the obtained permeate volume and the average flux decreased with the increasing number of enzyme reuse cycles (Fig. 5a). Third, the 3′-sialyllactose concentration in the permeate stayed above 1.5 mM during all the enzyme reuse cycles (Fig. 5b), indicating that Tr6 was quite stable and mechanically resistant even when exposed to a very high shear stress (i.e. agitation during filtration and high speed centrifugation after the third cycle). Our previous study has also shown that the enzyme retained 100% activity for 100 min even at 45 °C (Zeuner et al., 2014). After four cycles, there was only 0.5 mM 3′-sialyllactose in the retentate but more than 1.0 mM sialic acid, which was produced by the hydrolysis of both 3′-sialyllactose and CGMP in the presence of Tr6 (Zeuner et al., 2014). Thus, *in situ* separation of the product from the reactor resulted not only in reuse of the enzyme but also reduced product hydrolysis. Fourth, the centrifugation process did not obviously increase the permeate flux in the fourth cycle (Fig. 5a) because CGMP aggregation was negligible at 4 °C even at a high CGMP concentration of 240 g L^{-1}. The lack of CGMP aggregation can be explained by the fact that CGMP is negatively charged at pH 6.5 and electrostatic repulsion between the molecules hindered the aggregation or gelation of CGMP. Farías et al. (2010) found that the gelation time for a 80–100 g L^{-1} CGMP solution (25 °C) at pH 4 was 25 h while it decreased to several hours at pH 3.5. Therefore, it was possible to form and remove CGMP aggregates or gels after a long time standing and/or at a low pH. Otherwise, elevating the applied pressure to increase the CGMP concentration in the retentate might be an alternative to facilitate the deposition of CGMP. As long as the accumulation of CGMP is avoided, the batch EMR process could work continuously with reused enzymes and the biocatalytic productivity of Tr6 would largely increase.

3.3. Separation of 3′-sialyllactose and lactose by NF membranes (NF1)

As mentioned previously, an excess of lactose was used in the reaction in order to promote 3′-sialyllactose formation. Therefore sialyllactose, sialic acid, and lactose were found in the permeate from the EMR. Technically, the 3′-sialyllactose can be purified by anion exchange chromatography and subsequent lyophilization (Michalak et al., 2014). However, an extremely high concentration of lactose (>100 mM) could result in serious fouling of the chromatographic column and furthermore, the eluent (ammonium formate) used in chromatography could make the reuse of lactose more difficult. Nanofitration (NF), a membrane separation technology using both electric charge (Donnan effect) and pore size (sieving effect), could be used to separate charged and neutral solutes with low molecular weights (Luo and Wan, 2013). Sialyllactose is negatively charged while lactose is neutral, and furthermore, lactose has a lower molecular weight (Table 1). NF is therefore a promising candidate for the separation of 3′-sialyllactose and lactose. This speculation was confirmed by the results shown in Fig. 6. For the ETNA01PP, NP010 and NTR7450 membranes, the retention of 3′-sialyllactose was much higher than the retention of lactose. However, for the NF45 membrane with low molecular weight cut-off (MWCO), the retention of both solutes was 100% and the permeate flux was the lowest of all the tested membranes. The ETNA01PP membrane had highest flux but lowest retention of 3′-sialyllactose, which would result in a huge loss of product and therefore the ETNA01PP membrane could not be used for NF1. Although both membranes have the same MWCO, the retention of 3′-sialyllactose was much higher for the NP010 membrane than the ETNA01PP membrane, implying that the charge effect was more significant for the NP010 membrane due to its PES surface

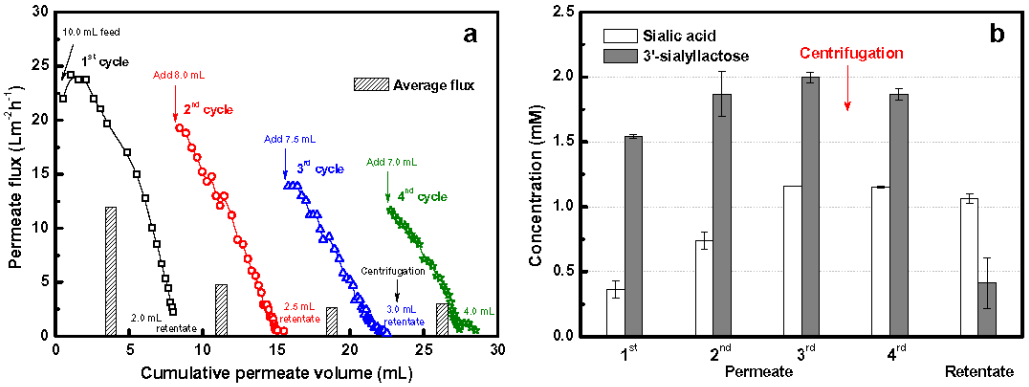


Fig. 5. (a) Permeate flux and (b) product concentration during four cycles in a free enzymatic membrane reactor (EMR) with PLCC membrane (UF2 for recycling enzyme and removing product). 20 min reaction time without pressure at 50 rpm for each cycle; filtrations were carried out at 4.0 bar and 250 rpm; new substrate was added to the reactor to keep the same initial feed volume (10 mL); after the third cycle, the retentate was centrifuged at 4 °C for 15 min (10,000×g) and the supernatant was returned to the reactor; after the fourth cycle, the retentate was analyzed by HPLC.

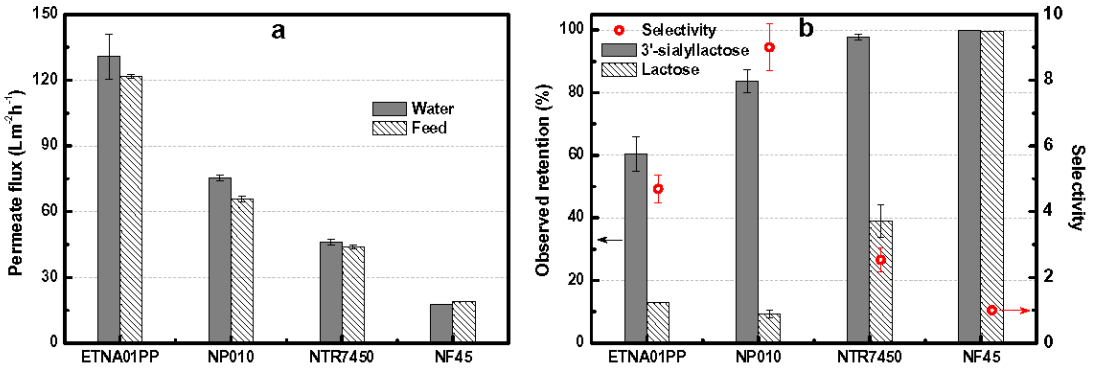


Fig. 6. (a) Average permeate flux and (b) observed retention and selectivity of solutes for different NF membranes. TMP = 4 bar; agitation speed = 250 rpm; 3′-sialyllactose = 0.5 mM; lactose = 0.5 mM; pH = 7.

layer. The NTR7450 membrane had an extremely high retention of 3′-sialyllactose (98%) due to the stronger charge and sieving effect, however, its lactose retention was not low enough (40%), resulting in a much lower selectivity than for the NP010 membrane. Therefore, the NP010 membrane with high flux and selectivity of solutes seems to be a suitable choice for the separation of 3′-sialyllactose and lactose (NF1). Despite having lower flux and selectivity, the NTR7450 membrane is also competitive thanks to its high retention of 3′-sialyllactose (Fig. 6).

In order to remove the lactose from the product obtained from the EMR (UF2), diafiltration operation was required in this step (NF1), and the NP010 and NTR7450 membranes were selected to evaluate the diafiltration process. A simple model based on a transient mass balance and the retention equation was utilized for this rapid evaluation, assuming that constant volume diafiltration (CVD) was employed and solute retention was constant during the diafiltration (Foley, 2006):

$$C_R = C_f \cdot \exp\left[(R_{obs} - 1) \cdot \frac{V_W}{V_f}\right] \quad (6)$$

where C_R and C_f are the solute concentrations in the retentate and initial feed, respectively; V_w is the water volume consumed in diafiltration and V_f is the initial feed volume.

Concentrations of 3′-sialyllactose and lactose in the feed were set at 4 and 113 mM (90% molar yield of 3′-sialyllactose). The predicted results according to Eq. (6) are shown in Table 3. With increasing water consumption, more lactose passed through the membrane and thus the purity of 3′-sialyllactose increased. The results from the NP010 and NTR7450 membranes were compared in terms of the purity and loss of 3′-sialyllactose at different water consumptions (Table 3). Although the product purity was higher with the NP010 membrane at each value of V_w/V_f, the loss of 3′-sialyllactose was too large when this membrane was used (62.5% 3′-sialyllactose loss when the purity was 75.3%). When the NTR7450 membrane was used, 97.4% of lactose was predicted to be removed and the loss of 3′-sialyllactose was only 12.4% when V_w/V_f was set as 6. Thus, it was obvious that the NTR7450 outperformed the NP010 membrane during the diafiltration. The experimental results from NTR7450 membrane also showed that only 11.3% of 3′-sialyllactose was washed away and 99.3% of lactose was removed, indicating that Eq. (6) was feasible to predict the solute concentration during the diafiltration though there was a small deviation for the prediction of lactose concentration (see Table 3, the rejection of lactose might change a little with the concentrations from 0.5 to 113 mM). After diafiltration, most of lactose in the product was taken away and the retained 3′-sialyllactose could be further concentrated and then purified by anion exchange chromatography.

3.4. Concentration of lactose by NF membranes (NF2)

The resulting permeate from the NF diafiltration (NF1) contains diluted lactose, which needs to be concentrated for reuse. Fig. 6b shows that the NF45 membrane is a suitable candidate for the concentration of the permeate from NF1 due to its high retention of lactose (99.6%, see Fig. 6b). Moreover, the concentration of lactose by NF has been widely and thoroughly studied in previous works (Cuartas-Uribe et al., 2007; Li et al., 2008; Luo et al., 2011). The performance of five NF membranes for lactose concentration was compared by Luo et al. (2011), revealing that the NF90 (Dow-Filmtec) membrane had the highest lactose retention, i.e. 99.3% while the NF270 (Dow-Filmtec) membrane was considered to be the best choice because it had the highest permeate flux and an acceptable lactose retention (97.8%). Therefore, the concentrate from NF2 could be returned into the EMR and the permeate (water) may

Table 3
Predicted results for the diafiltration process based on retention equation and mass balance under constant volume diafiltration mode.[a]

V_w/V_f	C_R of 3′-sialyllactose (mM)		C_R of lactose (mM)		Purity of 3′-sialyllactose (%)		Loss of 3′-sialyllactose (%)	
	NP010	NTR7450	NP010	NTR7450	NP010	NTR7450	NP010	NTR7450
0	4.00	4.00	113	113	3.42	3.42	0	0
1	3.40	3.91	45.7	61.4	6.92	5.99	15.1	2.18
2	2.89	3.83	18.5	33.4	13.5	10.3	27.9	4.30
3	2.45	3.74	7.46	18.1	24.7	17.1	38.7	6.39
4	2.08	3.66	3.01	9.85	40.8	27.1	48.0	8.42
5	1.77	3.58	1.22	5.35	59.2	40.1	55.8	10.4
6	1.50	3.51	0.49	2.91	75.3	54.7	62.5	12.4

[a] R_{obs} data were taken from Fig. 6b. The final concentrations of 3′sialyllacctose and lactose from experimental results for NTR7450 were 3.55 and 0.74 mM, respectively, indicating that this model is feasible.

be used for diafiltration in NF1, which could make the large-scale production of 3′-sialyllactose by Tr6 more economically feasible.

4. Conclusions

This work demonstrated that an integrated UF/NF membrane system for biocatalytic production of 3′-sialyllactose is a viable method for the valorization of dairy by-products using an engineered sialidase. The PLCC membrane is suitable to pretreat CGMP, separate the 3′-sialyllactose/lactose mixture from the EMR and reuse the enzyme because of its excellent antifouling performance, high CGMP retention and high permeation of 3′-sialyllactose. The permeate flux in the EMR decreased due to the accumulation of CGMP. The NTR7450 membrane was selected to isolate 3′-sialyllactose from lactose by diafiltration, and the diluted lactose could be concentrated by a tight NF membrane for reuse.

Acknowledgements

The authors would like to thank The Hans Christian Ørsted Postdoc Program (DTU) for financial support (J. Luo) and PhD student Yao Guo for the help with HPLC analysis. We would also like to thank Master students Rasmus A. Prior and Mads F. Andersen for the help with NF operations.

References

Abd El-Salam, M.H., 2006. Separation of casein glycomacropeptide from whey: methods of potential industrial application. Int. J. Dairy Sci. 1, 93–99.
Abels, C., Thimm, K., Wulfhorst, H., Spiess, A.C., Wessling, M., 2013. Membrane-based recovery of glucose from enzymatic hydrolysis of ionic liquid pretreated cellulose. Bioresour. Technol. 149, 58–64.
Chan, R., Chen, V., Bucknall, M.P., 2002. Ultrafiltration of protein mixtures: measurement of apparent critical flux, rejection performance, and identification of protein deposition. Desalination 146, 83–90.
Cuartas-Uribe, B., Vincent-Vela, M.C., Álvarez-Blanco, S., Alcaina-Miranda, M.I., Soriano-Costa, E., 2007. Nanofiltration of sweet whey and prediction of lactose retention as a function of permeate flux using the Kedem–Spiegler and Donnan steric partioning models. Sep. Purif. Technol. 56, 38–46.
Farías, M.E., Martinez, M.J., Pilosof, A.M.R., 2010. Casein glycomacropeptide pH-dependent self-assembly and cold gelation. Int. Dairy J. 20, 79–88.
Foley, G., 2006. Water usage in variable volume diafiltration: comparison with ultrafiltration and constant volume diafiltration. Desalination 196, 160–163.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvand, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: Walker, J.M. (Ed.), The Proteomics Protocol Handbook. Humana Press, Totowa, NJ, pp. 281–304.
Gilbert, M., Bayer, R., Cunningham, A.M., DeFrees, S., Gao, Y., Watson, D.C., Young, N.M., Wakarchuk, W.W., 1998. The synthesis of sialylated oligosaccharides using a CMP-Neu5Ac synthetase/sialyltransferase fusion. Nat. Biotechnol. 16, 769–772.
Giorno, F., Mazzei, R., Giorno, L., 2013. Purification of triacylglycerols for biodiesel production from Nannochloropsis microalgae by membrane technology. Bioresour. Technol. 140, 172–178.
Gopal, P.K., Gill, H., 2000. Oligosaccharides and glycoconjugates in bovine milk and colostrum. Br. J. Nutr. 84, S69–S74.
Hickey, R.M., 2012. The role of oligosaccharides from human milk and other sources in prevention of pathogen adhesion. Int. Dairy J. 22, 141–146.
Holck, J., Larsen, D.M., Michalak, M., Li, H., Kjærulff, L., Kirpekar, F., Gotfredsen, C.H., Forssten, S., Ouwehand, A.C., Mikkelsen, J.D., Meyer, A.S., 2014. Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by *Trypanosoma cruzi* trans-sialidase. New Biotechnol. 31, 156–165.
Leberknight, J., Wielenga, B., Lee-Jewett, A., Menkhaus, T.J., 2011. Recovery of high value protein from a corn ethanol process by ultrafiltration and an exploration of the associated membrane fouling. J. Membr. Sci. 366, 405–412.
Li, Y., Shahbazi, A., Williams, K., Wan, C., 2008. Separate and concentrate lactic acid using combination of nanofiltration and reverse osmosis membranes. In: Adney, W., McMillan, J., Mielenz, J., Klasson, K.T. (Eds.), Biotechnology for Fuels and Chemicals. Humana Press, pp. 369–377.
Luo, J., Wan, Y., 2013. Effects of pH and salt on nanofiltration – a critical review. J. Membr. Sci. 438, 18–28.
Luo, J., Ding, L., Qi, B., Jaffrin, M.Y., Wan, Y., 2011. A two-stage ultrafiltration and nanofiltration process for recycling dairy wastewater. Bioresour. Technol. 102, 7437–7442.
Martinez-Ferez, A., Rudloff, S., Guadix, A., Henkel, C.A., Pohlentz, G., Boza, J.J., Guadix, E.M., Kunz, C., 2006. Goat's milk as a natural source of lactose-derived oligosaccharides: isolation by membrane technology. Int. Dairy J. 16, 173–181.
Michalak, M., Larsen, D.M., Jers, C., Almeida, J.R.M., Willer, M., Li, H., Kirpekar, F., Kjærulff, L., Gotfredsen, C.H., Nordvang, R.T., Meyer, A.S., Mikkelsen, J.D., 2014. Biocatalytic production of 3′-sialyllactose by use of a modified sialidase with superior trans-sialidase activity. Process Biochem. 49, 265–270.
Neelima, Sharma, R., Rajput, Y.S., Mann, B., 2013. Chemical and functional properties of glycomacropeptide (GMP) and its role in the detection of cheese whey adulteration in milk: a review. Dairy Sci. Technol. 93, 21–43.
Paris, G., Ratier, L., Amaya, M.F., Nguyen, T., Alzari, P.M., Frasch, A.C.C., 2005. A sialidase mutant displaying trans-sialidase activity. J. Mol. Biol. 345, 923–934.
Puro, L., Kallioinen, M., Mänttäri, M., Natarajan, G., Cameron, D.C., Nyström, M., 2010. Performance of RC and PES ultrafiltration membranes in filtration of pulp mill process waters. Desalination 264, 249–255.
Qi, B., Luo, J., Chen, G., Chen, X., Wan, Y., 2012. Application of ultrafiltration and nanofiltration for recycling cellulase and concentrating glucose from enzymatic hydrolyzate of steam exploded wheat straw. Bioresour. Technol. 104, 466–472.
Silva-Hernandez, E.R., Nakano, T., Ozimek, L., 2002. Isolation and analysis of κ-casein glycomacropeptide from goat sweet whey. J. Agric. Food Chem. 50, 2034–2038.
Urashima, T., Taufik, E., Fukuda, K., Asakuma, S., 2013. Recent advances in studies on milk oligosaccharides of cows and other domestic farm animals. Biosci. Biotechnol. Biochem. 77, 455–466.
Zeuner, B., Luo, J., Nyffenegger, C., Aumala, V., Mikkelsen, J.D., Meyer, A.S., 2014. Optimizing the biocatalytic productivity of an engineered sialidase from *Trypanosoma rangeli* for 3′-sialyllactose production. Enzyme Microb. Technol. 55, 85–93.

Chapter 7: Overall conclusion and perspectives

7.1 Enzyme discovery

A total of six novel wild type enzymes were analyzed in the PhD project of which three (one trans-sialidase and two β-N-acetylhexosaminidases with trans-hexosaminidase activity) were found to have relevance to HMO production. The novel trans-sialidase (of *Haemophilus parasuis*) was identified based on an *in silico* 3D prediction approach where an aromatic sandwich was identified as a trans-glycosylation marker. The successful prediction is believed to be a result of a correct assessment, that the aromatic sandwich found in trypanosomal trans-sialidases could have (did) evolve (presumably in parallel) in a different species. Furthermore, the additional enzymes, expressed and evaluated in the trans-sialidase identification study, reiterate the (in literature) repeated suggestion that the aromatic sandwich is crucial to trans-sialidase activity in the TcTS. Further analysis of the novel trans-sialidase should be carried out to investigate whether additional traits of this interesting enzyme can be used to identify more trans-sialidases. The proof of concept on the aromatic sandwich as a trans-sialidase marker should result in establishment of more research studies where refined algorithms are used to screen for sialidases with aromatic sandwiches in the correct constellation. It should also be investigated if aromatic sandwiches can be found in other GH families and whether such enzymes have enhanced trans-glycosylation activity.

A classical enzyme discovery approach, in the form of metagenomic library screening, was used for the successful identification of the two novel β-hexosaminidases. The identified β-hexosaminidases were capable of carrying out trans-glycosylation using both natural and synthetic GlcNAc donor substrates (whereas only a synthetic GalNAc donor substrate was tested). Thus precursors for HMOs were produced in the form of LNT2, using the new enzymes in enzymatic reactions.

7.2 Enzyme development

A total of 8 mutants of which 5 have not previously been synthesized were constructed, expressed and evaluated for trans-sialidase activity. Integrating a range of mutations suggested by different research groups led to the development of two novel TrSA mutants with heavily improved trans-sialidase activity. To set this achievement in perspective it was evaluated that the enzymes are now at a state where it may be possible to sell the rights to use them for industrial application.

However, identification of common structural features between the trypanosomal trans-sialidases and the newly identified trans-sialidase of *H. parasuis* should be carried out, to identify further possibilities for improvements. It is therefore advised that a model of the novel trans-sialidase is produced based on a crystal structure, since the created homology model may be subject to inconsistencies with the actual structure.

7.3 Reaction and Process optimization

In the BioEng research group product purification was prior to this study only approached with academic use of the products in mind. During this PhD project however and integrated UF and NF membrane system was developed for the production of 3'SL. The process is believed to be industrially viable since several concerns such as product purity, sanitation, and substrate economy is catered to. By application of the integrated membrane system two major improvements were achieved:

- Improved substrate economy due to reduced loss of CGMP in preprocessing, enabled CGMP recovery, and lactose recovery for reuse in synthesis.
- Improved biocatalytic productivity, since product can be removed without compromising the enzyme.

7.4 Evaluation of PhD project goals

In all areas where HMO production was researched, in this PhD project, scientific as well as applicational progress was made. In the past chapter the scientific work has been presented and it clear that all goals of the PhD project was achieved.

References

A, W. J. (1985). ALTERATION OF SUBSTRATE SPECIFICITY OF SUBTILISIN BY CASSETTE MUTAGENESIS. *Journal of Cellular Biochemistry Supplement*, (9 PART B).

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143–69. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239358&tool=pmcentrez&rendertype=abstract

Amaya, M. ., Buschiazzo, A., Nguyen, T., & Alzari, P. M. (2003). The high resolution structures of free and inhibitor-bound Trypanosoma rangeli sialidase and its comparison with T. cruzi trans-sialidase. *J.Mol.Biol.*, *325*, 773–784. http://doi.org/12507479

Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., … Alzari, P. M. (2004). Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase. *Structure (London, England : 1993)*, *12*(5), 775–84. http://doi.org/10.1016/j.str.2004.02.036

Arbia, W. (2013). Chitin Extraction from Crustacean Shells Using Biological Methods - A Review. *Food Technology and Biotechnology*, *51*(1), 12 – 25.

Autenrieth, C., & Ghosh, R. (2015). Random mutagenesis and overexpression of rhodopin-3,4-desaturase allows the production of highly conjugated carotenoids in Rhodospirillum rubrum. *Archives of Biochemistry and Biophysics*, *572*(Sp. Iss. SI), 134–141. http://doi.org/10.1016/j.abb.2015.01.023

Badran, A. H., & Liu, D. R. (2015). In vivo continuous directed evolution. *Current Opinion in Chemical Biology*, *24*, 1–10. http://doi.org/10.1016/j.cbpa.2014.09.040

Ban, L., Pettit, N., Li, L., Stuparu, A. D., Cai, L., Chen, W., … Mrksich, M. (2012). Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry. *Nature Chemical Biology*, *8*(9), 769–73. http://doi.org/10.1038/nchembio.1022

Bech, L., Busk, P. K., & Lange, L. (2015). Cell wall degrading enzymes in Trichoderma asperellum grown on wheat bran. *Fungal Genomics & Biology*. Retrieved from http://vbn.aau.dk/en/publications/cell-wall-degrading-enzymes-in-trichoderma-asperellum-grown-on-wheat-bran(4443a042-64b3-4258-b26e-74f5b3ceb697).html

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*, *22*(9), 1147–62. http://doi.org/10.1093/glycob/cws074

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, *382*(Pt 3), 769–81. http://doi.org/10.1042/BJ20040892

Bornet, F. R. J., Brouns, F., Tashiro, Y., & Duvillier, V. (2002). Nutritional aspects of short-chain fructooligosaccharides: natural occurrence, chemistry, physiology and health implications. *Digestive and Liver Disease*, *34*, S111–S120. http://doi.org/10.1016/S1590-8658(02)80177-3

Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*, *485*(7397), 185–94. http://doi.org/10.1038/nature11117

Brody, E. P. (2000). Biological activities of bovine glycomacropeptide. *The British Journal of Nutrition*, *84 Suppl 1*, S39–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11242445

Brown, G. R., Hem, V., Katz, K. S., Ovetsky, M., Wallin, C., Ermolaeva, O., … Murphy, T. D. (2015). Gene: a gene-centered information resource at NCBI. *Nucleic Acids Research*, *43*(Database issue), D36–42. http://doi.org/10.1093/nar/gku1055

Buscaglia, C. A., Campo, V. A., Frasch, A. C. C., & Di Noia, J. M. (2006). Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nature Reviews. Microbiology*, *4*(3), 229–36. http://doi.org/10.1038/nrmicro1351

Busk, P. K., & Lange, L. (2013). Function-based classification of carbohydrate-active enzymes by recognition of short, conserved peptide motifs. *Applied and Environmental Microbiology*, *79*(11), 3380–91. http://doi.org/10.1128/AEM.03803-12

Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome Research*, *2*(1), 28–33. http://doi.org/10.1101/gr.2.1.28

Charalampopoulos, D., & Rastall, R. A. (2012). Prebiotics in foods. *Current Opinion in Biotechnology*, *23*(2), 187–91. http://doi.org/10.1016/j.copbio.2011.12.028

Daddaoua, A., Puerta, V., Zarzuelo, A., Suárez, M. D., Sánchez de Medina, F., & Martínez-Augustin, O. (2005). Bovine glycomacropeptide is anti-inflammatory in rats with hapten-induced colitis. *The Journal of Nutrition*, *135*(5), 1164–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15867298

Eigel, W. N., Butler, J. E., Ernstrom, C. A., Farrell, H. M., Harwalkar, V. R., Jenness, R., & Whitney, R. M. (1984). Nomenclature of Proteins of Cow's Milk: Fifth Revision. *Journal of Dairy Science*, *67*(8), 1599–1631. http://doi.org/10.3168/jds.S0022-0302(84)81485-X

Espinosa-Martos, I., & Rupérez, P. Soybean oligosaccharides. Potential as new ingredients in functional food. *Nutrición Hospitalaria*, *21*(1), 92–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16562819

F, H. (1983). WHEY RE CYCLING OF A WASTE PRODUCT. *Dirasat Natural Sciences (Amman)*, *10*(1), 57 – 66.

Fitzpatrick, K. (2012). WHEY PROTEIN: Advances in Nutrition and Applications. *Prepared Foods*, *181*(1).

Francis, J. C., & Hansche, P. E. (1972). Directed evolution of metabolic pathways in microbial populations. I. Modification of the acid phosphatase pH optimum in S. cerevisiae. *Genetics*, *70*(1), 59–73. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1212723&tool=pmcentrez&rendertype=abstract

Franck, A. (2002). Technological functionality of inulin and oligofructose. *The British Journal of Nutrition*, *87 Suppl 2*, S287–91. http://doi.org/10.1079/BJNBJN/2002550

Gavlighi, H. A., Meyer, A. S., & Mikkelsen, J. D. (2013). Enhanced enzymatic cellulose degradation by cellobiohydrolases via product removal. *Biotechnology Letters*, *35*(2), 205–12. http://doi.org/10.1007/s10529-012-1067-4

Ghazi, S. (2014). UV mutagenesis for the overproduction of xylanase from Bacillus mojavensis PTCC 1723 and optimization of the production condition. *Iranian Journal of Basic Medical Sciences*, *17*(11), 844 – 853.

Ghorbel-Bellaaj, O., Younes, I., Maâlej, H., Hajji, S., & Nasri, M. (2012). Chitin extraction from shrimp shell waste using Bacillus bacteria. *International Journal of Biological Macromolecules*, *51*(5), 1196–201. http://doi.org/10.1016/j.ijbiomac.2012.08.034

Gibson, G. R., Probert, H. M., Loo, J. Van, Rastall, R. A., & Roberfroid, M. B. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutrition Research Reviews*, *17*(2), 259–75. http://doi.org/10.1079/NRR200479

Gibson, G. R., & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics. *Journal of Nutrition*. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0029013322&partnerID=tZOtx3y1

Gilis, D. (2000). PoPMuSiC, an algorithm for predicting protein mutant stability changes. Application to prion proteins. *Protein Engineering*, *13*(12), 849 – 856.

Gross, M. (2011). Riding the wave of biological data. *Current Biology : CB*, *21*(6), R204–6. http://doi.org/10.1016/j.cub.2011.03.009

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews : MMBR*, *68*(4), 669–85. http://doi.org/10.1128/MMBR.68.4.669-685.2004

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal*, *280 (Pt 2*, 309–16. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1130547&tool=pmcentrez&rendertype=abstract

Holck, J., Hjernø, K., Lorentzen, A., Vigsnæs, L. K., Hemmingsen, L., Licht, T. R., … Meyer, A. S. (2011). Tailored enzymatic production of oligosaccharides from sugar beet pectin and evidence of differential effects of a single DP chain length difference on human faecal microbiota composition after in vitro fermentation. *Process Biochemistry*, *46*(5), 1039–1049. http://doi.org/10.1016/j.procbio.2011.01.013

Holck, J., Larsen, D. M., Michalak, M., Li, H., Kjærulff, L., Kirpekar, F., … Meyer, A. S. (2014). Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by Trypanosoma cruzi trans-sialidase. *New Biotechnology*, *31*(2), 156–65. http://doi.org/10.1016/j.nbt.2013.11.006

Jers, C., Michalak, M., Larsen, D. M., Kepp, K. P., Li, H., Guo, Y., … Mikkelsen, J. D. (2014). Rational design of a new Trypanosoma rangeli trans-sialidase for efficient sialylation of glycans. *PloS One*, *9*(1), e83902. http://doi.org/10.1371/journal.pone.0083902

Kahn, S. D. (2011). On the future of genomic data. *Science (New York, N.Y.)*, *331*(6018), 728–9. http://doi.org/10.1126/science.1197891

Kent, G. (2015). Global infant formula: monitoring and regulating the impacts to protect human health. *International Breastfeeding Journal*, *10*(1), 6. http://doi.org/10.1186/s13006-014-0020-7

Kihara, D. (Ed.). (2011). *Protein Function Prediction for Omics Era*. Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-0881-5

Kinch, L. N., & Grishin, N. V. (2002). Evolution of protein structures and functions. *Current Opinion in Structural Biology*, *12*(3), 400–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12127461

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(6), 524–36. http://doi.org/10.1002/bies.200900181

Koksharov, M. I., & Ugarova, N. N. (2008). Random mutagenesis of Luciola mingrelica firefly luciferase. Mutant enzymes with bioluminescence spectra showing low pH sensitivity. *Biochemistry (Moscow)*, *73*(8), 862–869. http://doi.org/10.1134/S0006297908080038

KOSHLAND, D. E. (1953). STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews*, *28*(4), 416–436. http://doi.org/10.1111/j.1469-185X.1953.tb01386.x

Kretz, K. A., Richardson, T. H., Gray, K. A., Robertson, D. E., Tan, X., & Short, J. M. (2004). Gene site saturation mutagenesis: a comprehensive mutagenesis approach. *Methods in Enzymology*, *388*, 3–11. http://doi.org/10.1016/S0076-6879(04)88001-7

Krishnan, V and Rupp, B. (2012). Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy. In *eLS*. John Wiley & Sons, Ltd: Chichester. http://doi.org/10.1002/9780470015902.a0002716.pub2

Kunz, C. (1999). Lactose-derived oligosaccharides in the milk of elephants: comparison with human milk. *BRITISH JOURNAL OF NUTRITION*, *82*(5), 391 – 399.

Kunz, C., & Rudloff, S. (1993). Biological functions of oligosaccharides in human milk. *Acta Paediatrica*, *82*(12), 903–912. http://doi.org/10.1111/j.1651-2227.1993.tb12597.x

Kunz, C., Rudloff, S., Baier, W., Klein, N., & Strobel, S. (2000). Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annual Review of Nutrition*, *20*, 699–722. http://doi.org/10.1146/annurev.nutr.20.1.699

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., … Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. http://doi.org/10.1038/35057062

Li, E. W. Y., & Mine, Y. (2004). Technical note: Comparison of chromatographic profile of glycomacropeptide from cheese whey isolated using different methods. *Journal of Dairy Science*, *87*(1), 174–7. http://doi.org/10.3168/jds.S0022-0302(04)73155-0

Loh, P.-R., Baym, M., & Berger, B. (2012). Compressive genomics. *Nature Biotechnology*, *30*(7), 627–30. http://doi.org/10.1038/nbt.2241

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014a). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), 490–495. http://doi.org/10.1093/nar/gkt1178

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014b). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(Database issue), D490–5. http://doi.org/10.1093/nar/gkt1178

Lorenz, P., & Eck, J. (2005). Metagenomics and industrial applications. *Nature Reviews. Microbiology*, *3*(6), 510–6. http://doi.org/10.1038/nrmicro1161

M, L. P. (1988). A NEW METHOD FOR RANDOM MUTAGENESIS OF COMPLETE GENES ENZYMATIC GENERATION OF MUTANT LIBRARIES IN-VITRO. *Protein Engineering*, *2*(1), 63 – 68.

Michalak, M., Larsen, D. M., Jers, C., Almeida, J. R. M., Willer, M., Li, H., … Mikkelsen, J. D. (2014). Biocatalytic production of 3'-sialyllactose by use of a modified sialidase with superior trans-sialidase activity. *Process Biochemistry*, *49*(2), 265–270. http://doi.org/10.1016/j.procbio.2013.10.023

Miyazaki, K., & Arnold, F. H. (1999). Exploring Nonnatural Evolutionary Pathways by Saturation Mutagenesis: Rapid Improvement of Protein Function. *Journal of Molecular Evolution*, *49*(6), 716–720. http://doi.org/10.1007/PL00006593

Montagna, G., Cremona, M. L., Paris, G., Amaya, M. F., Buschiazzo, A., Alzari, P. M., & Frasch, A. C. C. (2002). The trans-sialidase from the african trypanosome Trypanosoma brucei. *European Journal of Biochemistry / FEBS*, *269*(12), 2941–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12071958

Moore, J. C., & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology*, *14*(4), 458–467. http://doi.org/10.1038/nbt0496-458

Moshfegh, A. (1999). Presence of inulin and oligofructose in the diets of Americans. *Journal of Nutrition*, *129*(7).

Murata, T., Inukai, T., Suzuki, M., Yamagishi, M., & Usui, T. (n.d.). Facile enzymatic conversion of lactose into lacto-N-tetraose and lacto-N-neotetraose. *Glycoconjugate Journal*, *16*(3), 189–195. http://doi.org/10.1023/A:1007020219275

Myers, R. M., Lerman, L. S., & Maniatis, T. (1985). A general method for saturation mutagenesis of cloned DNA fragments. *Science*, *229*(4710), 242–247. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0022390906&partnerID=tZOtx3y1

Naumoff, D. G. (2011). Hierarchical classification of glycoside hydrolases. *Biochemistry. Biokhimiia*, *76*(6), 622–35. http://doi.org/10.1134/S0006297911060022

Nielsen, M., Lundegaard, C., Lund, O., & Petersen, T. N. (2010). CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Research*, *38*(Web Server issue), W576–81. http://doi.org/10.1093/nar/gkq535

Ninonuevo, M. R., Park, Y., Yin, H., Zhang, J., Ward, R. E., Clowers, B. H., … Lebrilla, C. B. (2006). A strategy for annotating the human milk glycome. *Journal of Agricultural and Food Chemistry*, *54*(20), 7471–80. http://doi.org/10.1021/jf0615810

Niyonzima, F. N., & More, S. S. (2015). Coproduction of detergent compatible bacterial enzymes and stain removal evaluation. *Journal of Basic Microbiology*, *55*(10), 1149–58. http://doi.org/10.1002/jobm.201500112

Ohnuma, S., Nakazawa, T., Hemmi, H., Hallberg, A. M., Koyama, T., Ogura, K., & Nishino, T. (1996). Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis. *The Journal of Biological Chemistry*, *271*(17), 10087–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8626566

Osorio, L., Ríos, I., Gutiérrez, B., & González, J. (2012). Virulence factors of Trypanosoma cruzi: who is who? *Microbes and Infection / Institut Pasteur*, *14*(15), 1390–402. http://doi.org/10.1016/j.micinf.2012.09.003

Paris, G., Ratier, L., Amaya, M. F., Nguyen, T., Alzari, P. M., & Frasch, A. C. C. (2005). A sialidase mutant displaying trans-sialidase activity. *Journal of Molecular Biology*, *345*(4), 923–34. http://doi.org/10.1016/j.jmb.2004.09.031

Patel, S., & Goyal, A. (2012). The current trends and future perspectives of prebiotics research: a review. *3 Biotech*, *2*(2), 115–125. http://doi.org/10.1007/s13205-012-0044-x

Pereira, M., Zhang, K., & Gong, Y. (1996). Invasive phenotype of Trypanosoma cruzi restricted to a population expressing trans-sialidase. *Infection and …*, *64*(9), 3884–3892.

Peter Tomme, R. Antony J. Warren, Robert C. Miller, Douglas G. Kilburn, & Neil R. Gilkes. (1996). *Enzymatic Degradation of Insoluble Carbohydrates*. (J. N. Saddler & M. H. Penner, Eds.) (Vol. 618). Washington, DC: American Chemical Society. http://doi.org/10.1021/bk-1995-0618

Pierdominici-Sottile, G., Palma, J., & Roitberg, A. E. (2014). Free-energy computations identify the mutations required to confer trans-sialidase activity into Trypanosoma rangeli sialidase. *Proteins*, *82*(3), 424–35. http://doi.org/10.1002/prot.24408

Polaina, J. (2007). *Industrial enzymes : structure, function and applications*. Springer.

Pontes-de-Carvalho, L. C., Tomlinson, S., & Nussenzweig, V. (1993). Trypanosoma rangeli sialidase lacks trans-sialidase activity. *Molecular and Biochemical Parasitology*, *62*(1), 19–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8114822

Prabu, K. (2012). Isolation and FTIR spectroscopy characterization of chitin from local sources. *Advances in Applied Science Research*, *3*(3).

Roberfroid, M. (2007). Prebiotics: the concept revisited. *The Journal of Nutrition*, *137*(3 Suppl 2), 830S–7S. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17311983

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., … Goodman, R. M. (2000). Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541–2547. http://doi.org/10.1128/AEM.66.6.2541-2547.2000

Sabater, C., Prodanov, M., Olano, A., Corzo, N., & Montilla, A. (2016). Quantification of prebiotics in commercial infant formulas. *Food Chemistry*, *194*, 6–11. http://doi.org/10.1016/j.foodchem.2015.07.127

Saito, T., Yamaji, A., & Itoh, T. (1991). A New Isolation Method of Caseinoglycopeptide from Sweet Cheese Whey. *Journal of Dairy Science*, *74*(9), 2831–2837. http://doi.org/10.3168/jds.S0022-0302(91)78463-4

Schnoes, A. M., Brown, S. D., Dodevski, I., & Babbitt, P. C. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Computational Biology*, *5*(12), e1000605. http://doi.org/10.1371/journal.pcbi.1000605

Scudder, P., Doom, J. P., Chuenkova, M., Manger, I. D., & Pereira, M. E. (1993). Enzymatic characterization of beta-D-galactoside alpha 2,3-trans-sialidase from Trypanosoma cruzi. *J. Biol. Chem.*, *268*(13), 9886–9891. Retrieved from http://www.jbc.org/content/268/13/9886.short

Sela, D. A., & Mills, D. A. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends in Microbiology*, *18*(7), 298–307. http://doi.org/10.1016/j.tim.2010.03.008

Shiian, S. D., Zueva, V. S., Nasonov, V. V, Zhigis, L. S., Tsoĭ, N. T., & Bovin, N. B. [Sialylation of N-carbohydrate chains of glycoproteins with immobilized trans-sialidase from Trypanosoma cruzi]. *Bioorganicheskaia Khimiia*, *30*(4), 400–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15469015

Smith, L. E., & Eichinger, D. (1997). Directed mutagenesis of the Trypanosoma cruzi trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. *Glycobiology*, *7*(3), 445–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9147054

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–8. http://doi.org/10.1093/nar/gki408

Taylor, C. M., & Woonton, B. W. (2009). Quantity and carbohydrate content of glycomacropeptide fractions isolated from raw and heat-treated milk. *International Dairy Journal*, *19*(12), 709–714. http://doi.org/10.1016/j.idairyj.2009.06.010

THIEM, J., & SAUERBREI, B. (2010). ChemInform Abstract: Chemoenzymatic Syntheses of Sialyloligosaccharides with Immobilized Sialidase. *ChemInform*, *23*(6), no–no. http://doi.org/10.1002/chin.199206271

Tiralongo, E., Martensen, I., Grötzinger, J., Tiralongo, J., & Schauer, R. (2003). Trans-sialidase-like sequences from Trypanosoma congolense conserve most of the critical active site residues found in other trans-sialidases. *Biological Chemistry*, *384*(8), 1203–13. http://doi.org/10.1515/BC.2003.133

Uchiyama, T., & Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*, *20*(6), 616–622. http://doi.org/10.1016/j.copbio.2009.09.010

Urashima, T., Taufik, E., Fukuda, K., & Asakuma, S. (2013). Recent advances in studies on milk oligosaccharides of cows and other domestic farm animals. *Bioscience, Biotechnology, and Biochemistry*, *77*(3), 455–66. http://doi.org/10.1271/bbb.120810

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, *291*(5507), 1304–51. http://doi.org/10.1126/science.1058040

Villumsen, N. S., Jensen, H. B., Thu Le, T. T., Møller, H. S., Nordvang, R. T., Nielsen, L. R., … Larsen, L. B. (2015). Self-assembly of caseinomacropeptide as a potential key mechanism in the formation of visible storage induced aggregates in acidic whey protein isolate dispersions. *International Dairy Journal*, *49*, 8–15. http://doi.org/10.1016/j.idairyj.2015.05.003

Wang, K., Lu, Y., Liang, W. Q., Wang, S. Di, Jiang, Y., Huang, R., & Liu, Y. H. (2012). Enzymatic Synthesis of Galacto-oligosaccharides in an Organic–Aqueous Biphasic System by a Novel β-Galactosidase from a Metagenomic Library. *Journal of Agricultural and Food Chemistry*, *60*(15), 3940–3946. http://doi.org/10.1021/jf300890d

Yu, G.-X., Park, B.-H., Chandramohan, P., Munavalli, R., Geist, A., & Samatova, N. F. (2005). In silico discovery of enzyme-substrate specificity-determining residue clusters. *Journal of Molecular Biology*, *352*(5), 1105–17. http://doi.org/10.1016/j.jmb.2005.08.008

Zaccolo, M., & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 β-lactamase. *Journal of Molecular Biology*, *285*(2), 775–783. http://doi.org/10.1006/jmbi.1998.2262

Zeuner, B., Jers, C., Mikkelsen, J. D., & Meyer, A. S. (2014). Methods for improving enzymatic trans-glycosylation for synthesis of human milk oligosaccharide biomimetics. *Journal of Agricultural and Food Chemistry*, *62*(40), 9615–31. http://doi.org/10.1021/jf502619p

Zeuner, B., Luo, J., Nyffenegger, C., Aumala, V., Mikkelsen, J. D., & Meyer, A. S. (2014). Optimizing the biocatalytic productivity of an engineered sialidase from Trypanosoma rangeli for 3'-sialyllactose production. *Enzyme and Microbial Technology*, *55*, 85–93. http://doi.org/10.1016/j.enzmictec.2013.12.009

Zhang, S.-B., & Wu, Z.-L. (2011). Identification of amino acid residues responsible for increased thermostability of feruloyl esterase A from Aspergillus niger using the PoPMuSiC algorithm. *Bioresource Technology*, *102*(2), 2093–2096. http://doi.org/10.1016/j.biortech.2010.08.019